

# Enhancing Race and Ethnicity Information in Medicaid Data: The Role of Census Bureau Data

SEHSD Working Paper 2023-05

Aubrey Limburg  
Adam Kurczewski,  
Victoria Udalova\*

February 2, 2023

## Abstract

Medicaid data are used frequently to investigate racial and ethnic disparities in health. However, there is considerable variation in the completeness of race/ethnicity information in Medicaid data across the United States (U.S.). To address these gaps, the U.S. Census Bureau's Enhancing Health Data (EHealth) Program assessed the feasibility, benefit, and effectiveness of linking Medicaid enrollment data (T-MSIS) with Census Bureau microdata (i.e., Decennial Census, American Community Survey (ACS), Census Numident) to enable researchers to document and track racial/ethnic disparities in health more effectively. Additionally, this research evaluated whether and to what degree bias was introduced into mortality estimates when Medicaid beneficiaries with missing race/ethnicity information were omitted from analysis. Although Decennial/ACS data reduced missing race/ethnicity information (from 19.47% to 7.11%), this varied considerably across states. Findings suggest that the exclusion of beneficiaries with missing race/ethnicity did not pose a problem for estimating race-specific mortality for most states, but researchers should be cautious about using Medicaid enrollment data for states with a high degree of bias. Results from this research demonstrate significant potential for using Census Bureau data to complement existing health records that commonly lack important demographic characteristics, such as race/ethnicity. Overall, we find that enhancing race/ethnicity information in Medicaid data with restricted Census Bureau microdata is feasible and can advance our understanding of population health.

---

Limburg and Kurczewski are survey statisticians at the U.S. Census Bureau. Udalova is a Senior Economist and the Enhancing Health Data (EHealth) Program manager at the Census Bureau. This report is released to inform interested parties of ongoing research and to encourage discussion. Any opinions and conclusions expressed herein are those of the authors and do not reflect the views of the Census Bureau. This DMS number for this project is P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy. We would like to thank numerous internal and external reviewers who shared their expertise and valuable feedback.

\* Corresponding author: Victoria Udalova, [victoria.m.udalova@census.gov](mailto:victoria.m.udalova@census.gov).

## Table of Contents

Introduction.....	1
Background.....	2
Racial and Ethnic Disparities in Health.....	2
Missing Race/Ethnicity Information in Medicaid Data.....	2
Concordance of Race/Ethnicity Information Across Data Sources.....	4
Current Study.....	4
Data and Methods.....	5
Data.....	5
Measures.....	7
Analytic Approach.....	9
Results.....	10
Data Linkage.....	10
Race/Ethnicity Concordance.....	16
Bias Assessment.....	21
Conclusions.....	29
References.....	32
Supplemental Materials.....	39
Appendix A. Data Linkage Supplement.....	39
Appendix B. Race/Ethnicity Concordance Supplement.....	50
Appendix C. Bias Assessment Supplement.....	68

## Introduction

Medicaid claims and enrollment data have been used frequently as a source for investigating longstanding racial and ethnic disparities in health and health care. However, there is a high level of missing race/ethnicity data among Medicaid beneficiaries (Ng et al. 2017) and this varies considerably across states (Fernandez et al. 2015; Khan et al. 2019). For example, in 2016, seven states reported missing race/ethnicity information for 50% or more of their beneficiaries (Centers for Medicare & Medicaid Services 2022). If race/ethnicity information is not missing at random, then the incompleteness of the data may have significant implications for the study of racial and ethnic disparities in health (Brown-Podgorski, Roberts, and Schpero 2022; MACPAC 2021b).

The problem of missing race/ethnicity information is not unique to Medicaid claims and enrollment data and is common in other types of health records (Branham et al. 2022; Klinger et al. 2015; Krieger et al. 2020; Polubriaginof et al. 2019). Therefore, it is imperative to identify novel solutions to enhance the quality of Medicaid data, which may help advance our understanding of the root causes of racial and ethnic disparities in health among this particularly at-risk population and to develop effective solutions to close these gaps.

As a means of evaluating and enhancing existing Medicaid data, this study linked Medicaid enrollment data from 2016 with a variety of restricted U.S. Census Bureau microdata sources (i.e., Decennial Census and American Community Survey (ACS) data). Additionally, we assessed inconsistencies in race/ethnicity responses for beneficiaries between Medicaid and Decennial/ACS data and implemented strategies to create congruence. Finally, we tested a use case, race-specific all-cause mortality, to assess whether using data with missing race/ethnicity information resulted in biased estimates of racial/ethnic disparities in health.

Our paper is not the first to link Medicaid enrollment data to restricted Census Bureau microdata. Previous research linked 2006-2008 Medicaid enrollment data to Decennial and ACS data to evaluate race/ethnicity concordance (Fernandez et al. 2015). We build on and expand this work by examining contemporary data that has undergone several important changes since 2008. First, we examined the period following the Patient Protection and Affordable Care Act (ACA), which expanded Medicaid eligibility in many states and altered the characteristics of the Medicaid population. The ACA may have resulted in changes to the racial/ethnic composition of Medicaid beneficiaries across states and over time. Second, we examined Medicaid data in the new format (T-MSIS rather than MSIS) which resulted in significant changes to data collection and reporting, which occurred between 2014 and 2016. Additionally, and most importantly, our study explicitly evaluated potential bias introduced into the estimates of health outcomes and health disparities due to missing race/ethnicity information in Medicaid data.

Overall, this research demonstrates that linking Medicaid and Census Bureau microdata sources is a feasible and effective tool for filling in missing race/ethnicity information for Medicaid beneficiaries and can help to advance our understanding of population health. We find that Census Bureau microdata have significant potential to complement existing health records that commonly lack important demographic characteristics, such as race and ethnicity. Additionally, health records may aid the Census Bureau in identification of hard-to-count populations as well as in

validation and enhancement of Census Bureau data to provide richer data on the social determinants of health. The data used in this paper were restricted microdata that could be used only for approved Census Bureau research projects (we provide more detail about this in the data section). Despite the restricted nature of these data, we hope that this research can serve as a guide to those interested in using Medicaid data for research on health disparities by demonstrating when missing race/ethnicity data for beneficiaries pose challenges to the estimation of racial and ethnic disparities in health.

## **Background**

### *Racial and Ethnic Disparities in Health*

A large and growing body of work has documented extensive racial and ethnic disparities in both health and health care in the United States (U.S.). For example, racial and ethnic minority individuals are less likely to be insured relative to White, non-Hispanic individuals (Smedley, Stith, and Nelson 2003; Kirby and Kaneda 2010). Such disparities in health insurance coverage may account for some differences in health care access based on race (Lillie-Blanton and Hoffman 2005). However, even when people are insured, disparities in health persist due to lower quality of care, residential segregation, and structural and systemic racism (Smedley, Stith, and Nelson 2003). Identifying ways to reduce these disparities has been a major research and policy priority for several decades (U.S. Department of Health and Human Services 2020). However, little progress has been made (Mahajan et al. 2021) and the coronavirus disease 2019 (COVID-19) pandemic has further laid bare these disparities (Polyakova et al. 2021).

To date, research has documented important racial/ethnic disparities among Medicaid beneficiaries across a variety of different health outcomes (e.g., asthma, diabetes, pregnancy outcomes) and service utilization outcomes (e.g., substance abuse services, HPV vaccination) (Buescher, Whitmire, and Pullen-Smith 2010; Heflinger, Chatman, and Saunders 2006; Silber et al. 2017; Staras et al. 2010; Zhang et al. 2013). Medicaid expansion under the ACA has enhanced access to health insurance (Baumgartner et al. 2020), but research has continued to demonstrate the persistence of racial/ethnic disparities in health and health care outcomes (Lee and Porell 2020). Therefore, despite having identical health insurance, racial and ethnic disparities persist even in the Medicaid context.

Medicaid enrollment data is a vital source for understanding health disparities since Medicaid is a federally supported and mandated program that serves primarily low-income individuals. In 2019, Medicaid provided health insurance to over a quarter of the U.S. population and two-fifths of beneficiaries identified as having a race or ethnicity other than White or non-Hispanic (MACPAC 2021a). Although Medicaid data are available to researchers, they are currently missing race/ethnicity information at varying rates across states in the U.S., posing potential problems to research validity (Centers for Medicare & Medicaid Services 2022; Khan et al. 2019).

### *Missing Race/Ethnicity Information in Medicaid Data*

At the time of writing this paper, the Centers for Medicare & Medicaid Services (CMS) did not require states to collect race and ethnicity information and there was considerable state-level

variation in missing race/ethnicity data (Centers for Medicare & Medicaid Services 2022; Khan et al. 2019).<sup>1</sup> This variation is in part due to the optional status of race/ethnicity reporting on Medicaid applications as well as differences in state-level collection of this information (Zylla and Lukanen 2021).

The Data Quality Atlas (DQ Atlas), maintained by CMS, provides state-level information for a variety of Medicaid variables and is updated annually to provide information about Medicaid data quality (Centers for Medicare & Medicaid Services 2022). DQ Atlas provides information based on state-specific Medicaid race/ethnicity distributions as well as public ACS data. In 2016 (release 2), just under one-third of all states (n=16) were deemed “low concern” regarding missing race/ethnicity information. Low concern was classified as 10% or fewer of beneficiaries missing race/ethnicity. On the other hand, race/ethnicity data in seven states were deemed unusable, meaning that more than 50% of beneficiaries were missing this information. The remaining 28 states were considered either medium or high concern in terms of the level of missingness of race/ethnicity information among Medicaid beneficiaries. The overall distribution of states in each of the race/ethnicity data quality categories has changed little over the years, suggesting a longer-term problem beyond just 2016 data (Melendez et al. 2022).<sup>2</sup>

Despite the challenge of missing race/ethnicity data, a large body of work has relied on Medicaid data to investigate racial and ethnic disparities in health and health care. Such studies have attempted to circumvent issues related to missing race and ethnicity data. First, some research has omitted Medicaid beneficiaries from analytic samples when they are missing race and/or ethnicity information (Bilaver, Sobotka, and Mandell 2021; McConnell et al. 2018; Staras et al. 2010). Second, rather than omitting observations with missing race/ethnicity, some research has included Medicaid beneficiaries with missing race/ethnicity information into an other race or non-White race category (Davis et al. 2017; Marton et al. 2016; Samnaliev, McGovern, and Clark 2009; Tangka et al. 2017). Although this approach increases statistical power, these analyses may result in biased estimates of race-specific health outcomes. Third, rather than including all racial/ethnic groups in the analytic sample, some studies have focused exclusively on a limited number of racial/ethnic identities in their analyses (Buescher, Whitmire, and Pullen-Smith 2010; Cantor et al. 2020; Fabius et al. 2018; Heflinger, Chatman, and Saunders 2006; Horvitz-Lennon et al. 2015; Ringwalt et al. 2015; Silber et al. 2017; Zhang et al. 2013). In this last approach, it is unclear how observations with unknown race/ethnicity information were handled. Despite attempts to mitigate the challenges of missing Medicaid data elements, these strategies have limitations and their effects on the studies’ results and conclusions remain unknown.

Given the extent and variation of missing race/ethnicity data among Medicaid beneficiaries, it is challenging to accurately document, track, and inform interventions to reduce health disparities. Lack of access to high quality data on race/ethnicity could lead to biased and misleading estimates. More specifically, if race/ethnicity information was missing completely at random then we would

---

<sup>1</sup> States are only required to collect information deemed necessary for making a determination regarding program eligibility, of which race/ethnicity is not deemed necessary (42 CFR § 435.907).

<sup>2</sup> As of 2020, 16 states were considered “low concern” while four states had race/ethnicity data that was considerable unusable,

expect estimates of disparities in health outcomes to be unbiased even when not able to account for missing race/ethnicity information. It is more likely, however, that race/ethnicity information is not missing at random given previous research demonstrating significant differences between those with and without race/ethnicity information (Fernandez et al. 2015; Kressin et al. 2003; Ennis et al. 2018). As a result, by excluding individuals missing race/ethnicity, estimates of various health disparities may be biased. This bias could result in a difference in magnitude or even a change in the direction of the relationship.

### *Concordance of Race/Ethnicity Information Across Data Sources*

Beyond missing race/ethnicity information, an important component of linking Medicaid and Census Bureau data is that race/ethnicity available in both data sources may not be concordant for a given individual. There are a couple of reasons why race/ethnicity information may be discordant. First, differences in how questions regarding race and ethnicity are asked can impact responses (Compton et al. 2013; Miyawaki 2016; Terry and Fond 2013). Second, race and ethnic identity are subject to change over time (Liebler et al. 2017; Saperstein and Penner 2012). Additionally, previous research found that there are important differences in the characteristics of individuals whose race/ethnicity information is concordant across data sources relative to those whose race/ethnicity information varies across data sources (Ennis et al. 2018; Fernandez et al. 2015; Kressin et al. 2003; McAlpine et al. 2007).

Some strategies have been proposed to resolve discordant race/ethnicity information (Ennis et al. 2018). In their evaluation of how best to overcome discordant race and ethnicity responses, the authors found that for ethnicity (Hispanic origin) and race it was best to prioritize responses for smaller racial groups over responses for larger racial groups based on the 2010 Census population distribution. Previous research has yet to assess potential bias introduced to findings by omission or misclassification of those with missing race/ethnicity information within Medicaid enrollment data. Resolving discordant race/ethnicity information in addition to filling in missing information would provide the opportunity to assess potential bias introduced into estimates of health outcomes and disparities.

### *Current Study*

This study expands previous research in several important ways. First, we reinvestigate the feasibility and effectiveness of linking Medicaid enrollment data with Census Bureau microdata given a different composition of the Medicaid population following Medicaid expansion post-ACA and the implementation of a new Medicaid data collection system (T-MSIS). Second, we evaluate beneficiaries with discordant race/ethnicity information in Medicaid and Decennial/ACS data and employ procedures for creating race/ethnicity concordance. Finally, we empirically assess whether bias is introduced into estimates of race-specific all-cause mortality due to missing race/ethnicity information in Medicaid data for all U.S. states. Our findings have important implications for research focused on health disparities and help to assess whether research using Medicaid data has been misestimating racial and ethnic disparities in mortality.

## Data and Methods

### *Data*

This research relied on several data sources. First, we used Transformed Medicaid Statistical Information System (T-MSIS) Analytic Files (TAF) (referred to as TAF throughout the paper) for year 2016 (Release 2) for all 50 U.S. states and the District of Columbia (DC).<sup>3</sup> These data are collected individually from each state and include all Medicaid and Children’s Health Insurance Program (CHIP) beneficiaries for 2016. Data from 2016 were used since Medicaid data underwent a massive transformation between 2014 and 2016 when CMS transitioned from the MSIS to T-MSIS data collection system. Analysis of only 2016 data allowed for a five-and-a-half-year follow-up period to measure mortality.

Second, race/ethnicity data was derived from restricted Census Bureau microdata sources including Decennial Census 2000 and 2010 as well as American Community Survey (ACS) 2001-2019 (Decennial/ACS data, hereafter)<sup>4</sup>. Only the most recently reported race and ethnicity information was retained for a single person.

The linkage between TAF and Decennial/ACS data was performed using a unique individual-level anonymized identifier common across all data sources at the Census Bureau, called a Protected Identification Key (PIK). PIKs are created using personally identifiable information (PII) and are based on deterministic and probabilistic record linkage methods using the Personal Identification Validation System (PVS) process at the Census Bureau (Wagner and Layne 2014). Receiving a PIK depends on the quality of an individual’s PII and the likelihood of their information being found in various sources of administrative records data used during the PIK assignment process.<sup>5</sup> Therefore, not everyone receives a PIK, and young people, immigrants, and racial/ethnic minorities are less likely to receive a PIK (Bond et al. 2014; Rastogi and O’Hara 2012). After the linkage process, all TAF records were retained, regardless of whether they were linked to Decennial/ACS data.

Finally, TAF records were linked with the Census Bureau’s Numerical Identification file (Census Numident) to measure all-cause mortality among Medicaid beneficiaries (2021 quarter 2 vintage). The Census Numident covers all individuals with a Social Security Number (SSN) in the U.S. and provides a date of death for most deceased individuals (Finlay and Genadek 2021).

The data used in this paper were restricted microdata made available with anonymized linkage keys that could be used only for approved research projects. Analyses were conducted on Census Bureau restricted servers and no microdata ever left the server. Only aggregate output, which was reviewed and approved by the Census Bureau’s Disclosure Review Board, was released. External

---

<sup>3</sup> The TAF data dictionary can be found here: <https://resdac.org/cms-data/files/taf-de/data-documentation>.

<sup>4</sup> Imputed values were used, so every individual in the combined Decennial/ACS dataset had race and ethnicity information. Additional information regarding imputation of race and ethnicity information at the Census Bureau can be found here: <https://www.census.gov/newsroom/blogs/random-samplings/2021/08/census-when-demographic-and-housing-characteristics-are-missing.html>.

<sup>5</sup> PII is removed from the data prior to any researchers accessing the data. We only had access to the data that contain anonymized linkage keys and no PII.

researchers can request access to these restricted data by submitting their research proposal and obtaining a Special Sworn Status (SSS).<sup>6</sup> Since this is a costly process both in terms of time and resources, it may not be a feasible option for all researchers. Therefore, while our main analyses are succinct, we provide an abundance of information in the Supplemental Materials section.

Not all beneficiaries in TAF data were retained for analysis. First, about 2% of beneficiaries in TAF were not eligible for Medicaid in any month of year 2016 and were omitted from the analysis.<sup>7</sup> Second, Medicaid beneficiaries could have multiple observations in TAF data within a given year (Centers for Medicare & Medicaid Services 2021; Whitney et al. 2020)<sup>8</sup> due to individuals moving across states (one observation for each state in which the enrollee received benefits) or non-adjacent enrollment periods during a given year within the same state (i.e., churn). Since our analyses were primarily focused on the state level, beneficiaries with multiple observations in different states were retained and assigned to their respective states. However, those beneficiaries who had multiple observations in 2016 within the same state were deduplicated so that there was only one observation for a given beneficiary in 2016 within a given state. This restriction meant that if an individual had multiple Medicaid enrollment spells in 2016 in a single state, we collapsed all these spells into one record for this person. We relied on a unique Medicaid identifier (BENE\_ID)<sup>9</sup> to identify these observations.

In some instances, demographic information, such as race/ethnicity or date of birth, was not concordant across multiple spells for the same beneficiary. For beneficiaries with observations in different states, the information was left intact, and no changes were made. For example, if an enrollee lived in Colorado and was reported to be White and then moved to Montana and reported to be Black, no changes were made to the observations, despite the discrepancy. Instead, both records were retained, and one contributed to the Colorado beneficiary estimates and the other contributed to the Montana beneficiary estimates. However, there were instances in which the same beneficiary appeared in a state more than once during 2016 and had discordant race/ethnicity information. For these individuals, we relied on the race/ethnicity prioritization scheme developed by Ennis et al. (2018) to resolve the discordant information and retained just one observation per person in a given state.<sup>10</sup>

Third, TAF data had some instances in which people with the same BENE\_ID (thought to be unique to a person over time and place) had different birth dates, and the difference had a

---

<sup>6</sup> For more information on how external researchers can request access to restricted Census Bureau microdata, please see: <https://www.census.gov/about/adrm/fsrdc.html>.

<sup>7</sup> Individuals coded as “1” on MISG\_ELGBLTY\_DATA\_IND were omitted from analysis. These individuals are included in TAF data because they have claims data but no eligibility information from their prior Medicaid enrollment year.

<sup>8</sup> Here we are looking exclusively at BENE\_ID which is created and assigned by the CMS as part of their Chronic Conditions Warehouse. This identifier is meant to identify beneficiaries across states, time, and other CMS data sources.

<sup>9</sup> BENE\_ID is created by CMS by relying on a combination of MSIS\_ID, STATE\_CD, Social Security Number (SSN), date of birth, sex, and other identifying information. Some Medicaid enrollees are missing BENE\_IDs and missingness of BENE\_IDs is not consistent across states. For example, California has a high rate of missingness of BENE\_IDs (Whitney et al. 2020).

<sup>10</sup> Additional information about how we used this approach is discussed in the race/ethnicity concordance section.

significant range. We were unable to determine in which instances this was an error, a representation of a parent and child relationship (newborns may initially be assigned the same beneficiary identifier as their parent), or, in fact, these were two distinct individuals. Observations that did not have concordant birth dates were omitted. Finally, individuals missing age in TAF data were omitted from analysis to maintain consistency across all analyses. The final number of observations included in analysis was 95.99 million with 2.87% of original observations being omitted. The unit of analysis was person-spell. While it is best to interpret our estimates as representing person-spell level estimates rather than person-level estimates, we note that since there is a small proportion of individuals moving across states or having non-adjacent Medicaid spells within the same year the difference between person-level and person-spell level counts is very small.

*Measures*

Race and ethnicity were coded differently between TAF and Census Bureau data sources. TAF provided race/ethnicity information in a single variable that combined race and ethnicity (Table 1). Decennial/ACS data contained separate race and ethnicity variables based on the Office of Management and Budget (OMB) standards (Office of Management and Budget 1997). We collapsed race and ethnicity variables in Decennial/ACS data to align with the race/ethnicity variable in TAF (Table 1). We relied on four primary race/ethnicity categories: Hispanic (any race); White, non-Hispanic (White, hereafter); Black, non-Hispanic (Black, hereafter); and other, non-Hispanic (other, hereafter).

**Table 1. Recode of Broad Race/Ethnicity Categories in Decennial/ACS and TAF Data**

<b>Decennial/ACS broad race/ethnicity categories</b>	<b>TAF broad race/ethnicity categories<sup>11</sup></b>
1 – Hispanic (any race)	7 – Hispanic, any race
2 – White, non-Hispanic	1 – White, non-Hispanic
3 – Black, non-Hispanic	2 – Black, non-Hispanic
4 – Other, non-Hispanic	3 – Asian, non-Hispanic 4 – American Indian or Alaska Native (AIAN), non-Hispanic 5 – Hawaiian/Pacific Islander, non-Hispanic 6 – Multiracial, non-Hispanic
9 – Unknown*	Null/missing – source value is missing or unknown

\*This category is not available in Census Bureau data because there are no observations with unknown race/ethnicity. We recoded observations in TAF data with missing race/ethnicity as unknown race/ethnicity (category = 9) so that every observation was assigned a code.

Although our main analyses are based on the broad race/ethnicity categorization, we provide some supplementary analyses that break down Medicaid beneficiaries by more detailed race/ethnicity. In general, it is challenging to find data, survey or administrative records, that can be broken into detailed racial/ethnic categories, and even when possible, sample sizes are often not sufficient for

<sup>11</sup> For all races that were non-Hispanic, they were either reported as non-Hispanic or Hispanic origin was not reported.

some racial/ethnic groups to conduct analyses. We present some of our findings by detailed race/ethnicity categories to highlight the value of population-level Census Bureau microdata sources for these kinds of analyses (Table 2).

**Table 2. Recode of Detailed Race/Ethnicity Categories in Decennial/ACS and TAF Data**

<b>Decennial/ACS detailed race/ethnicity categories</b>	<b>TAF detailed race/ethnicity categories</b>
1 – Hispanic	20 – Hispanic, any race
2 – White, non-Hispanic	1 – White, non-Hispanic
3 – Black, non-Hispanic	2 – Black, non-Hispanic
4 – Asian, non-Hispanic	4 – Asian Indian, non-Hispanic 5 – Chinese, non-Hispanic 6 – Filipino, non-Hispanic 7 – Japanese, non-Hispanic 8 – Korean, non-Hispanic 9 – Vietnamese, non-Hispanic 10 – Other Asian, non-Hispanic 11 – Asian Unknown, non-Hispanic 12 – Multi-Asian, non-Hispanic
5 – American Indian and Alaska Native, non-Hispanic	3 – American Indian or Alaskan Native (AIAN), non-Hispanic
6 – Native Hawaiian and Other Pacific Islander, non-Hispanic	13 – Native Hawaiian, non-Hispanic 14 – Guamanian or Chamorro, non-Hispanic 15 – Samoan, non-Hispanic or Hispanic not reported 16 – Other Pacific Islander, non-Hispanic 17 – Native Hawaiian or Other Pacific Islander Unknown, non-Hispanic 18 – Multi-Islander, non-Hispanic
7 – Two or more races, non-Hispanic	19 – Multi-racial, non-Hispanic
9 – Unknown*	Null/missing – source value is missing or unknown

\*This category is not available in Census Bureau data because there are no observations with unknown race/ethnicity. Observations in TAF data with missing race/ethnicity were recoded as unknown race/ethnicity (9) so that every observation was assigned a code.

To assess the PIK rate, we included measures of whether observations received a PIK. We also present metrics on the linkage rate. An observation in TAF data was considered linked if it had a PIK and was found in Decennial/ACS data. There are no instances in which someone would be linked to Decennial/ACS data and not have a PIK. Furthermore, a Medicaid beneficiary may have a valid PIK but may not be linked to Decennial/ACS data. If an observation was not linked to Decennial/ACS data, this meant that they did not have race/ethnicity information in these sources. However, all observations that were successfully linked did have race/ethnicity information in Decennial/ACS data.

To assess potential bias that may be introduced into estimates of health outcomes and disparities due to missing race/ethnicity information, we used a measure of all-cause mortality as a case study. We focused on mortality because we had access to reliable date of death information and because this is one of the most studied health outcomes in health disparities research. We used a date of death variable from the Census Numident to construct race-specific all-cause mortality rates. The study period was split into two distinct periods to account for the potential impact of the COVID-19 pandemic: 1) January 2016 and February 2020 (before the pandemic) and 2) between January 2016 and June 2021 (including the pandemic). For mortality estimates stratified by state, we calculated an annual all-cause mortality rate per 10,000 beneficiaries for each demographic group. For example, the pre-pandemic (January 2016 – February 2020) all-cause mortality rate for Black beneficiaries was calculated by dividing the total number of Black beneficiaries who died during this period by the number of Black beneficiaries who were alive at the beginning of January 2016 for each state.

Our analysis of mortality rate did not adjust for differences in age across racial/ethnic groups. Our focus was on examining the bias – the difference between mortality estimates based on data with and without missing race/ethnicity information – rather than estimating demographically-adjusted mortality rates per se. However, we provided a series of supplementary analyses that compared mortality rates across different age groups. For those analyses, age was derived from TAF data and was recoded as a three-category variable: less than 18, 18-64, and 65 years of age and older.

Table 3 provides descriptive statistics for our Medicaid population stratified by race/ethnicity, as reported in TAF data. White beneficiaries were the largest race/ethnicity group (35.48%), while other race was the smallest (6.18%). On average, beneficiaries were 28.15 years of age, while Hispanic beneficiaries were, on average, the youngest (24.89 years of age) and other race beneficiaries were on average the oldest (32.53 years of age). Medicaid beneficiaries were also more likely to be female (55.19%) with only slight variation across race/ethnicity groups in TAF.

**Table 3. Descriptive Statistics for Medicaid Population**

	<b>Age (mean)</b>	<b>Age (SD)</b>	<b>Female (%)</b>	<b>Overall (%)</b>
<b>TAF Race/Ethnicity</b>	(1)	(2)	(3)	(4)
<b>Hispanic</b>	24.89	20.71	55.97	22.00
<b>White</b>	31.07	23.08	55.34	35.48
<b>Black</b>	27.25	21.54	56.05	16.87
<b>Other</b>	32.53	23.97	54.87	6.18
<b>Unknown</b>	25.88	22.54	53.4	19.47
<b>Overall</b>	28.15	22.45	55.19	100

Source: T-MSIS Analytic Files (TAF) data (2016)

Notes: SD = standard deviation; DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

### *Analytic Approach*

The analysis was split into three parts: 1) data linkage, 2) race/ethnicity concordance, and 3) bias assessment. First, for the data linkage, we linked TAF and Decennial/ACS data and provided information regarding all Medicaid beneficiaries as well as those who had PIKs and those who

were linked to Decennial/ACS. We also provided national maps documenting the amount of missingness for race/ethnicity in TAF data across all states, including DC. Second, for the concordance component, we evaluated whether race/ethnicity responses were concordant between TAF and Decennial/ACS data. We also implemented two possible solutions for overcoming missingness and discordance of race/ethnicity information in TAF data, which we discuss below. Finally, for the bias assessment, we provided an analytic case that evaluated potential bias introduced into estimates of race-specific all-cause mortality when relying on TAF data with missing race/ethnicity. Race-specific all-cause mortality rates were calculated before and after the linkage to Decennial/ACS to identify how mortality rates changed for each state. Analyses presented in this paper are based on the Medicaid population in 2016 and hence, formal tests of statistical significance were not conducted. All results presented in this paper have gone through internal review within the Census Bureau and follow the Census Bureau's Disclosure Review Board guidelines for rounding and disclosure of information (DRB # CBDRB-FY22-375). Additionally, all numeric values presented were rounded according to Census Bureau disclosure protocols to preserve data privacy and therefore may not add to 100 percent due to rounding error.

## **Results**

### *Data Linkage*

Table 4 provides the race/ethnicity distribution of Medicaid beneficiaries for all states using information from TAF data and shows PIK rates and linkage rates to Decennial/ACS<sup>12</sup>. Overall, the sample consisted of 95.99 million observations. Of these observations, 19.47% had missing race/ethnicity information in 2016 TAF data (Column 2). Of all the TAF observations, 97.73% received a PIK (Column 4), and 69.85% were linked to Decennial/ACS (Column 6). The PIK rate was highest for White (99.53%) beneficiaries followed by Black (99.38%), other race (98.68%), and unknown race/ethnicity (97.59%) beneficiaries. The PIK rate was lowest for those who identified as Hispanic in TAF data (93.42%). Conditional on receiving a PIK, 71.47% of observations were found in Decennial/ACS (Column 7).

---

<sup>12</sup> Note that individuals born after April 1, 2010 did not have race/ethnicity information in the 2000 or 2010 Decennial Census. Therefore, young children in 2016 TAF data (<6 years of age) may only have race/ethnicity information in TAF or in ACS data.

**Table 4. Race/Ethnicity Distribution Based on 2016 TAF Data**

TAF Race/Ethnicity	All Observations		Observations with PIKs		Observations Linked to Decennial/ACS		
	N	Column %	N	PIK Rate (%)	N	Linkage Rate (%)	Linkage Rate (%), Conditional on PIK
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<b>Hispanic</b>	21,120,000	22.00	19,730,000	93.42	13,030,000	61.70	66.04
<b>White</b>	34,060,000	35.48	33,900,000	99.53	26,880,000	78.92	79.29
<b>Black</b>	16,190,000	16.87	16,090,000	99.38	11,380,000	70.29	70.73
<b>Other</b>	5,929,000	6.18	5,851,000	98.68	3,880,000	65.44	66.31
<b>Unknown</b>	18,690,000	19.47	18,240,000	97.59	11,890,000	63.62	65.19
<b>Total</b>	95,990,000	100	93,810,000	97.73	67,050,000	69.85	71.47

Source: T-MSIS Analytic Files (TAF) data (2016); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

The linkage rate between TAF and Decennial/ACS data was highest among White (78.92%), Black (70.29%), and other race (65.44%) beneficiaries and was lowest for Hispanic (61.70%) and unknown race/ethnicity (63.62%) beneficiaries (Column 6). Conditional on receiving a PIK, the linkage rate ranged from 65.19% for beneficiaries with unknown race/ethnicity to 79.29% for White beneficiaries (Column 7). The PIK rate was surprisingly high for beneficiaries with unknown race/ethnicity information in TAF data, meaning their PII was of sufficient quality and their information was captured in some administrative records used by the Census Bureau to assign the linkage keys. But, conditional on receiving a PIK, beneficiaries with unknown race/ethnicity had the lowest linkage rate to Decennial/ACS. This suggests that individuals who may be reluctant to self-report their race/ethnicity information do it consistently regardless of who is collecting this information.

Overall, by linking 2016 TAF and Decennial/ACS data it was possible to fill in race/ethnicity information for 63.62% of all the Medicaid beneficiaries missing this information. Even though Decennial/ACS data did not provide race/ethnicity information for every Medicaid beneficiary missing this information, this rate is high and indicates the potential value of Census Bureau data in supplementing race/ethnicity information in health records. Of the Medicaid beneficiaries with unknown race/ethnicity information in TAF data, we were unable to fill in race/ethnicity information for 36.38% of them and, thus, we were unable to assess how that group's racial/ethnic distribution may be different from the part of the unknown group that we were able to find in Decennial/ACS data. Similar analyses were also conducted for the detailed race/ethnicity variable (Table A-1). Since there are some Medicaid beneficiaries who were not assigned PIKs or, who despite having been assigned PIKs, were not found in Decennial/ACS data, we caution readers that our findings may not be generalizable to the entire Medicaid population.

Table 5 provides the race/ethnicity distribution for beneficiaries with and without race/ethnicity information in TAF data. Columns 1 and 2 provide the race/ethnicity distribution for beneficiaries

in TAF data who were not missing this information. Columns 3 and 4 provide the race/ethnicity distribution for beneficiaries who were missing this information in TAF but whose race/ethnicity information could be found in Decennial/ACS data.<sup>13</sup> This table does not include those who were missing race/ethnicity in TAF and who could not be linked to Decennial/ACS data. Overall, among beneficiaries with missing race/ethnicity in TAF who could be found in Decennial/ACS data, a smaller percentage identified as Hispanic while a larger percentage identified as White and other race (Column 4) relative to those not missing race/ethnicity in TAF (Column 2). The percentage of those identifying as Black was comparable between the two groups. We performed similar analyses comparing racial/ethnic distribution of beneficiaries with and without race/ethnicity information in TAF data for each state (Table A-2) as well as for the detailed race/ethnicity variable (Table A-3). State-level analysis shows how the racial/ethnic distribution of the Medicaid population changed in each state when missing race/ethnicity information was populated using Decennial/ACS data.

**Table 5. Comparing Race/Ethnicity Distribution for Beneficiaries with and without Race/Ethnicity Information in TAF Data**

	Not missing race in TAF (TAF reported race)		Missing TAF race (Decennial/ACS reported race)	
	N	Column %	N	Column %
	(1)	(2)	(3)	(4)
<b>Hispanic</b>	21,120,000	27.32	2,901,000	24.46
<b>White</b>	34,060,000	44.06	5,466,000	46.09
<b>Black</b>	16,190,000	20.94	2,435,000	20.53
<b>Other</b>	5,929,000	7.67	1,063,000	8.96
<b>Total</b>	77,300,000	100	11,860,000	100

Source: T-MSIS Analytic Files (TAF) data (2016); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

After linking TAF data with Decennial/ACS data, 7.11% of Medicaid beneficiaries nationally were still missing race/ethnicity (Table 6). Since race/ethnicity information was missing in both sources, these observations are not included in future analyses. There are a couple of reasons why Decennial/ACS data may not fill in missing information for every person. First, in order to link TAF data to any Census Bureau record, individuals must have valid PIKs which means they must first be found in an internal reference file consisting of records from many surveys, censuses, and administrative records – the main one of which is the Social Security Administration (SSA) administrative data. Second, we do not have race/ethnicity information for all observations with PIKs.

Among beneficiaries who were unable to be linked to Decennial/ACS data, the vast majority received a PIK. This finding is consistent with the notion that health records may capture populations who are not observed in any survey, census, or other administrative records data that

<sup>13</sup> Although, differences between the two groups in Table 5 may be representative of true differences, some of the difference could be due to differences in how race/ethnicity is collected.

the Census Bureau collects or houses.<sup>14</sup> Overall, most Medicaid beneficiaries (92.89%) had race/ethnicity information in one or both data sources with the majority (57.36%) having this information in both files.

**Table 6. Presence of Race/Ethnicity Information After Linking TAF and Decennial/ACS Data**

	<b>N</b>	<b>Column %</b>
	(1)	(2)
<b>Missing in both sources</b>	6,828,000	7.11
<b>Missing in TAF only</b>	11,860,000	12.36
<b>Missing in Decennial/ACS only</b>	22,230,000	23.16
<b>Have in both sources</b>	55,060,000	57.36
<b>Total</b>	95,990,000	100

Source: T-MSIS Analytic Files (TAF) data (2016); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

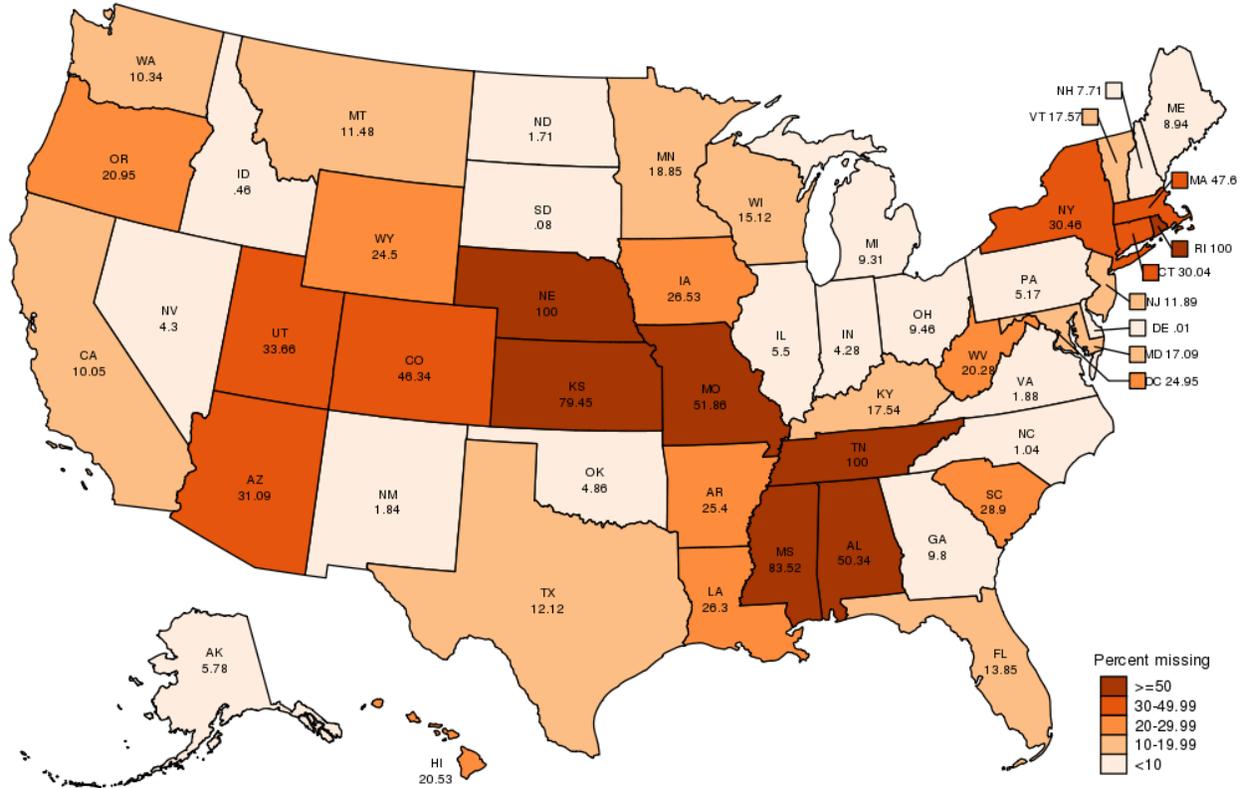
Notes: DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

Figure 1 shows the percent of beneficiaries who were missing race/ethnicity in 2016 TAF data across states (these values can also be found in Table A-5). Although, on average, 19.47% percent of Medicaid beneficiaries were missing race/ethnicity across all states, Figure 1 demonstrates that there was considerable variation across states. Some states did not have race/ethnicity information for any of their Medicaid beneficiaries, including Nebraska, Rhode Island, and Tennessee. Beyond states with complete missingness, there were also states with high rates of missingness including Mississippi (83.52%) and Kansas (79.45%). Additionally, the lowest rates of missingness were observed in Delaware (0.01%), South Dakota (0.08%), and Idaho (0.46%) suggesting that these states had race/ethnicity information for nearly all their Medicaid beneficiaries in 2016.

---

<sup>14</sup> Given the large amount of documentation required to establish eligibility for Medicaid, these data include individuals who are likely to be captured by survey, census, or other administrative records data that the Census Bureau collects or houses. Other health records, such as emergency department or hospital records, likely include more hard-to-reach populations and we would expect an even lower PIK rate for individuals captured by those records.

**Figure 1. Percent of Medicaid Beneficiaries with Missing Race/Ethnicity Information in 2016 TAF Data**

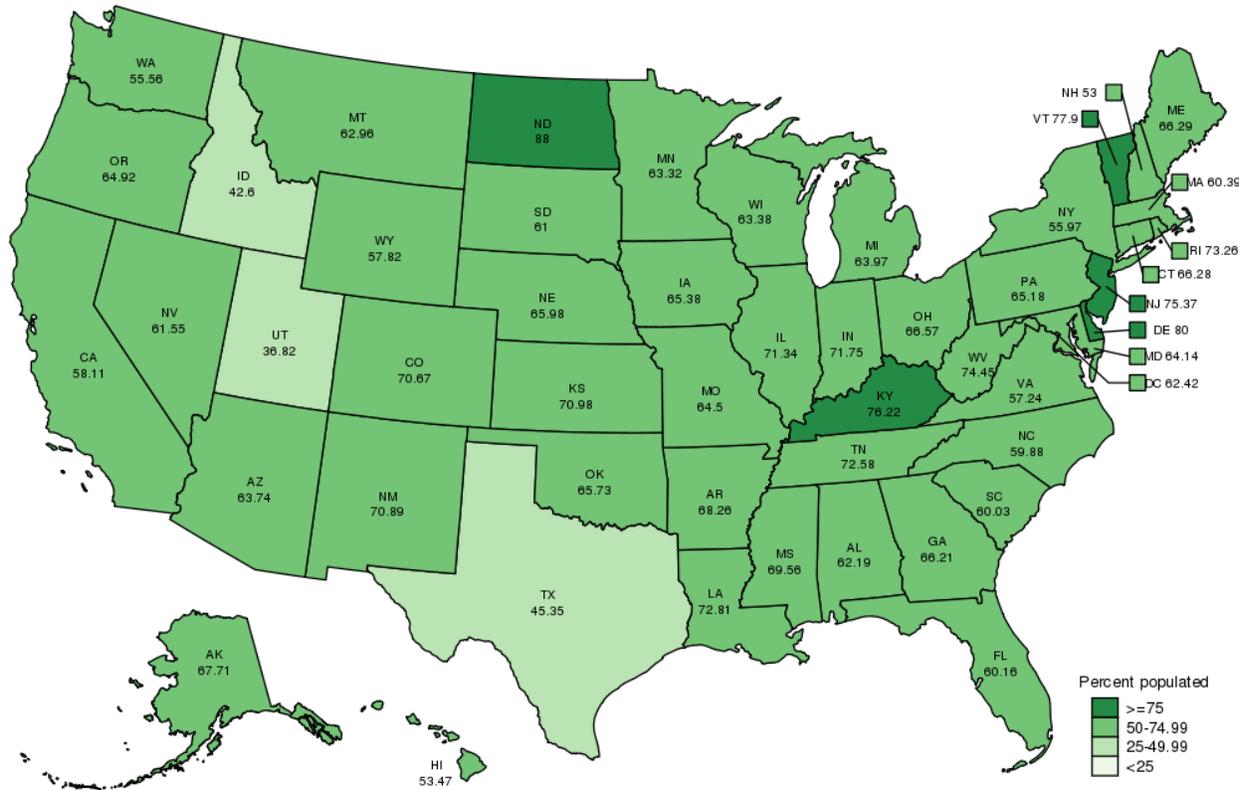


Source: T-MSIS Analytic Files (TAF) data (2016); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

Figure 2 depicts the percent of beneficiaries in TAF data who had missing race/ethnicity information but for whom we could obtain their race/ethnicity information from Decennial/ACS data, by state (these values can also be found in Table A-5). For example, of all the Medicaid beneficiaries missing race/ethnicity in Colorado in 2016 (46.34% in Figure 1), it was possible to populate race/ethnicity information for 70.67% of them by linking TAF to Decennial/ACS data (Figure 2). All states had at least one-third of their observations with missing race/ethnicity information populated and nearly one-tenth had over 75% of all missingness accounted for by Decennial/ACS data.

**Figure 2. Percent of Medicaid Beneficiaries in 2016 TAF Data with Missing Race/Ethnicity Information Populated by Decennial/ACS Data**

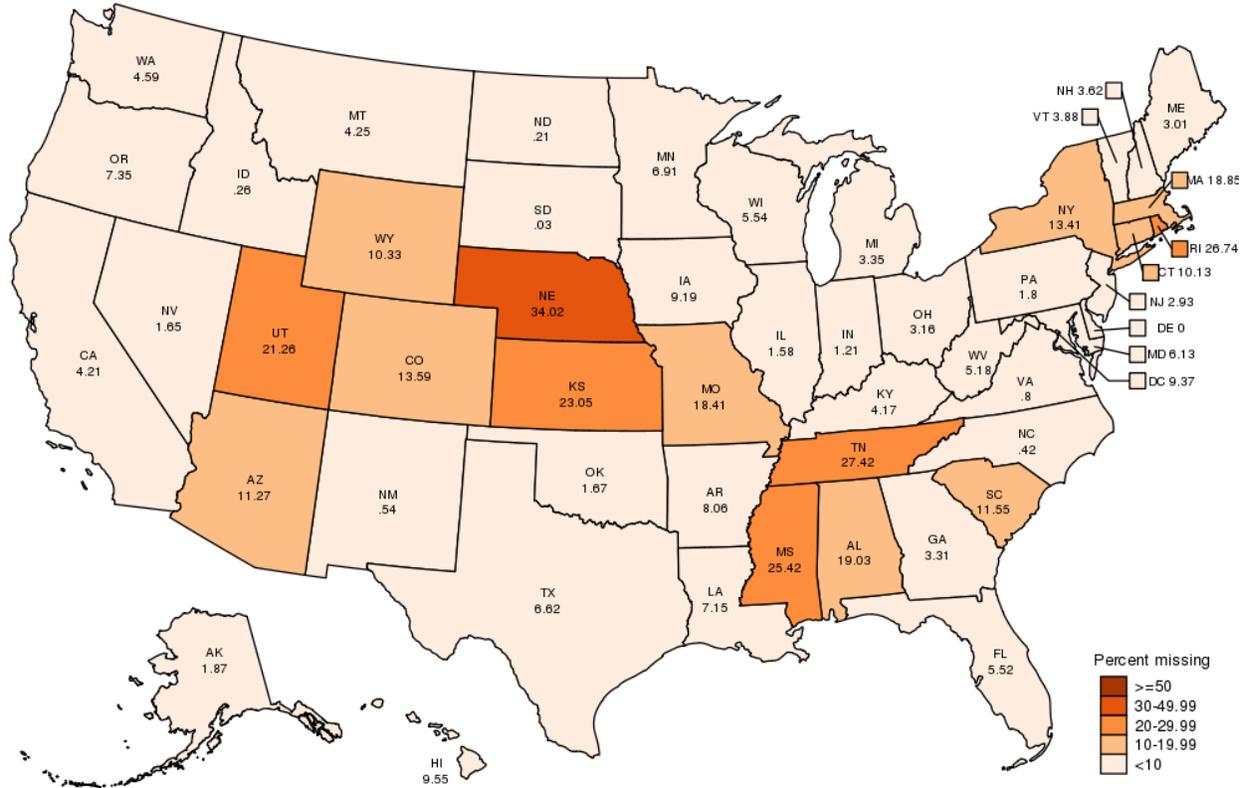


Source: T-MSIS Analytic Files (TAF) data (2016); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

Figure 3 depicts the percent of Medicaid beneficiaries still missing race/ethnicity information after 2016 TAF data was linked to Decennial/ACS data (these values can also be found in Table A-5). Relative to Figure 1, there is a marked decline in the percent of Medicaid beneficiaries missing race/ethnicity after race/ethnicity information was brought in from Decennial/ACS data. Across all states, the overall missing rate was reduced by 63.62%. Note that there were no longer any states with missing race/ethnicity information for 50% or more of their beneficiaries. After the linkage, most states (n=36) had 10 percent or fewer beneficiaries with missing race/ethnicity.

**Figure 3. Percent of Medicaid Beneficiaries with Missing Race/Ethnicity Information in 2016 TAF Data After Linking to Decennial/ACS Data**



Source: T-MSIS Analytic Files (TAF) data (2016); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

This first portion of this paper has demonstrated the feasibility and effectiveness of linking TAF to Decennial/ACS data. Overall, we found a high PIK rate (97.73%) and a moderate linkage rate (69.85%) for Medicaid beneficiaries. Additionally, we found that Decennial/ACS data could fill in 63.62% of observations with missing race/ethnicity information in 2016 TAF data. Overall, there were still 7.11% of TAF observations for which we could not identify race/ethnicity information in Decennial/ACS data. There were also important differences across states. There was considerable variation in both missingness rate as well as the effectiveness of linking to Decennial/ACS data across states.

### *Race/Ethnicity Concordance*

Our second goal was to evaluate whether race/ethnicity information was concordant between TAF Decennial/ACS data for linked observations. Based on previous research (Fernandez et al. 2015), we anticipated that the majority of beneficiaries would have concordant race/ethnicity information in both sources. Table 7 shows the concordance of race/ethnicity between the two data sources. Most individuals had race/ethnicity information in both data sources (57.36%). Of those with race/ethnicity information in both sources, 87.54% had concordant race/ethnicity information. For

the rest of the beneficiaries (12.47%), their race/ethnicity information was discordant between the two sources. In what follows, we focus on Medicaid beneficiaries with concordant race/ethnicity information and then we move our attention to Medicaid beneficiaries with discordant information.

**Table 7. Concordance of Race/Ethnicity Information in TAF and Decennial/ACS Data**

	<b>N</b>	<b>Column %</b>
	(1)	(2)
<b>Missing information in one or both sources</b>	40,930,000	42.64
<b>Has information in both sources</b>	55,060,000	57.36
<b>Discordant information</b>	6,865,000	12.47
<b>Concordant information</b>	48,200,000	87.54
<b>Total</b>	95,990,000	100

Source: T-MSIS Analytic Files (TAF) data (2016); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

In the Race/Ethnicity Concordance Supplement, we focus on observations with race/ethnicity concordance for each of the four broad race/ethnicity categories for all states (Figure B-1 through Figure B-4). Among those with race/ethnicity in both sources, the percent of Medicaid beneficiaries with concordant responses was defined as the number with concordant race/ethnicity between TAF and Decennial/ACS data divided by the number of beneficiaries reporting that race/ethnicity in TAF. Although both the overall PIK and linkage rates were high, evaluation of concordance rates demonstrated that there were differences in race/ethnicity concordance that varied by racial/ethnic group as well as by state. More specifically, race/ethnicity concordance was highest for beneficiaries who identified as Black (89.87%), White (88.08%), and Hispanic (86.16%) and lowest for beneficiaries who identified as other race (81.38%), with considerable variation across states.

For beneficiaries with discordant race/ethnicity information, we identified two approaches for creating concordance for two purposes. First, we wanted to further assess the potential value of race/ethnicity information in Decennial/ACS data for supplementing TAF data. Second, we wanted to investigate potential bias in estimations of health outcomes when missingness is unaccounted for.

In our first approach, we retained all TAF race/ethnicity information since this was directly reported by Medicaid beneficiaries to state Medicaid offices. For those who were missing race/ethnicity information in TAF data but were linked to Decennial/ACS data, we supplemented their race/ethnicity with information from Decennial/ACS data (from here on we will refer to the race/ethnicity variable constructed using this approach as “Supplemental Race”). However, the Supplemental Race strategy is limited in that it did not consider possible discrepancies for observations with discordant race/ethnicity information between TAF and Decennial/ACS data. Therefore, in the second approach, a new variable was constructed to address issues of discordance based on the methods proposed by Ennis et al. (2018). We created a new race variable in line with the authors suggested approach which prioritizes smaller racial/ethnic groups (from here on we

will refer to the race/ethnicity variable constructed using this approach as “Prioritized Race”). The Prioritized Race approach had one difference relative to Ennis et al. (2018). In our data, race/ethnicity was combined into one variable, while Ennis et al. (2018) treated race and ethnicity as discrete measures. Given this, our final prioritization scheme was as follows: Hispanic (any race); other race; Black; and White.<sup>15</sup> Our Prioritized Race approach had advantages over the Supplemental Race approach in that it overcame discordant race/ethnicity information and more fully utilized Decennial/ACS data. However, the way we achieved these extra benefits with this approach was by making additional assumptions about race/ethnicity prioritization as a means of resolving discordant information.

Table 8 provides the race/ethnicity distribution for the original TAF race variable (race/ethnicity as reported in TAF; referred to as TAF Race, hereafter) and our two approaches incorporating Decennial/ACS data: Supplemental Race (race/ethnicity as reported in TAF but supplemented by information from Decennial/ACS data for individuals with no TAF information) and Prioritized Race (prioritizing smaller racial/ethnic groups over larger ones when information from the two sources was discordant). When comparing the TAF Race and Supplemental Race, the racial/ethnic distributions were comparable. However, when comparing TAF Race and Prioritized Race, there was an increase in the percentage of Hispanic and other race beneficiaries and a decrease in the percentage of White beneficiaries. An increase in the percent of Hispanic beneficiaries for Prioritized Race highlights potential issues of incomplete and/or inconsistent collection of ethnicity data across state Medicaid offices. There was very little change in the percentage of Black beneficiaries. Overall, using Decennial/ACS data only to fill in race/ethnicity information for Medicaid beneficiaries missing this information in TAF data changed the racial/ethnic distribution minimally. On the other hand, using Decennial/ACS data to both fill in and override race/ethnicity information for certain individuals with discordant information, prioritizing smaller racial/ethnic groups, led to larger changes in the racial/ethnic distribution of the Medicaid population.

**Table 8. Race/Ethnicity Distribution in TAF Data Before and After Linking to Decennial/ACS Data**

	TAF Race		Supplemental Race		Prioritized Race	
	N	Column %	N	Column %	N	Column %
	(1)	(2)	(3)	(4)	(5)	(6)
<b>Hispanic</b>	21,120,000	27.32	24,020,000	26.94	26,580,000	29.81
<b>White</b>	34,060,000	44.06	39,530,000	44.34	36,330,000	40.75
<b>Black</b>	16,190,000	20.94	18,620,000	20.88	18,170,000	20.38
<b>Other</b>	5,929,000	7.67	6,992,000	7.84	8,085,000	9.07
<b>Total</b>	77,300,000	100	89,160,000	100	89,160,000	100

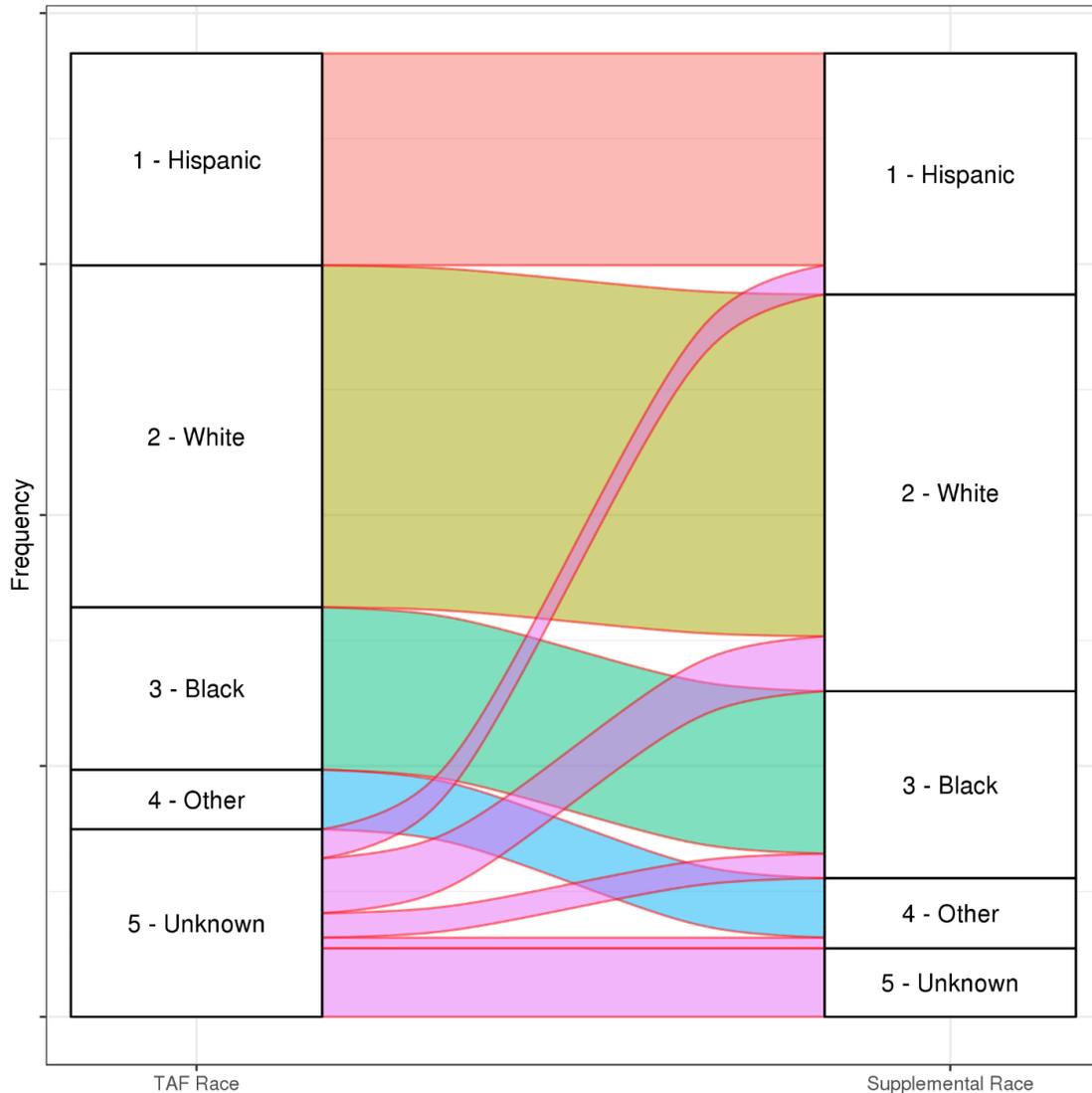
Source: T-MSIS Analytic Files (TAF) data (2016); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: This table omits the “unknown” race groups to create comparable race/ethnic distributions between the three approaches. DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

<sup>15</sup> For the detailed race/ethnicity variable, we prioritized race/ethnicity assignment in the following way: Hispanic (any race); Native Hawaiian Pacific Islander; American Indian and Alaskan Native; two or more races; Asian; Black; and White.

Figure 4 and Figure 5 demonstrate the flow of beneficiaries between TAF Race and the Supplemental/Prioritized Race approaches, respectively. In Figure 4, the only changes to the distribution were for beneficiaries with unknown race/ethnicity information. Of those beneficiaries for whom it was possible to assign race/ethnicity information, the majority were moved to the White race group based on Decennial/ACS data.

**Figure 4. Flow of Medicaid Beneficiaries from the TAF Race Approach to Supplemental Race Approach**

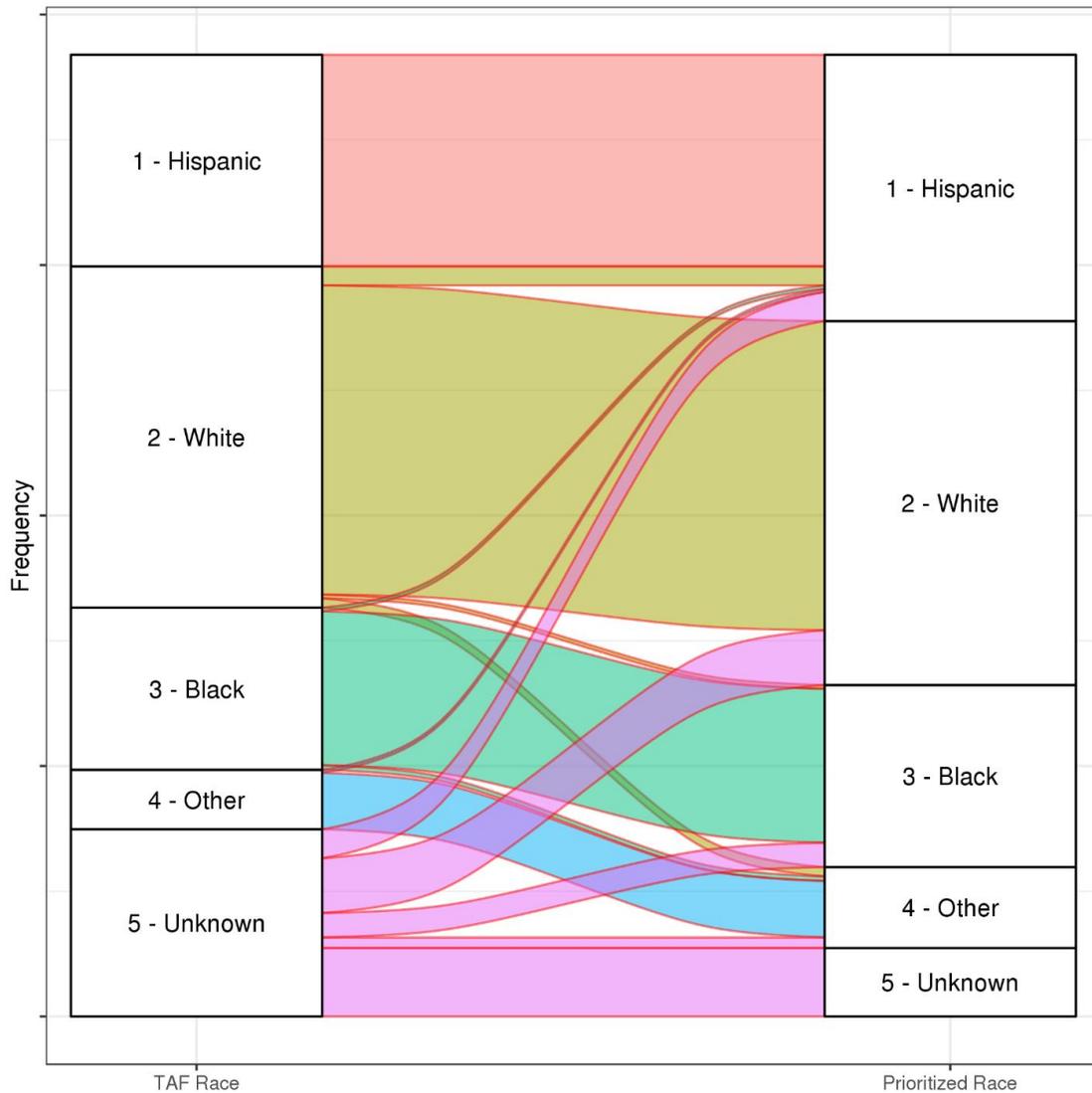


Source: T-MSIS Analytic Files (TAF) data (2016); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

More movement between race/ethnicity categories took place between TAF Race and Prioritized Race (Figure 5). All beneficiaries who reported a Hispanic identity in TAF Race retained that race/ethnicity identity in Prioritized Race. Besides those beneficiaries who had missing race/ethnicity information in TAF, no beneficiaries moved to the White race category based on the prioritization scheme. There was an increase in the overall number of individuals identifying as Hispanic, with movement from all race/ethnicity groups with the biggest contribution from the unknown race/ethnicity group in TAF Race. Among White Medicaid beneficiaries, a small proportion moved to Hispanic, Black, and other race based on the Prioritized Race approach.

**Figure 5. Flow of Medicaid Beneficiaries from the TAF Race Approach to the Prioritized Race Approach**



Source: T-MSIS Analytic Files (TAF) data (2016); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

We performed additional analyses examining state-level changes in the race/ethnicity distribution when comparing TAF Race to each of the two approaches – Supplemental and Prioritized Race (Figure B-5 through Figure B-12). The biggest benefit of linking TAF and Decennial/ACS data came from filling in race/ethnicity information for beneficiaries in states who were missing entire race/ethnicity categories (Connecticut, DC, Kansas) or all race/ethnicity information for all beneficiaries (Nebraska, Rhode Island, Tennessee). In general, changes across states for each race/ethnicity group were relatively small in magnitude. The greatest amount of absolute change in the race/ethnicity distribution was for White beneficiaries with a more pronounced change in the Prioritized Race approach.

### *Bias Assessment*

Our third and final aim was to assess whether using TAF data with missing race/ethnicity information results in biased estimates of racial/ethnic health disparities. Table 9 shows the average annual crude mortality rates per 10,000 Medicaid beneficiaries who were missing and those who were not missing race/ethnicity information in TAF data for the two periods. Columns 1 and 2 focus on the period before the pandemic (January 2016 – February 2020). Column 1 provides mortality rates using TAF Race for those not missing TAF Race. This is a group of Medicaid beneficiaries for whom researchers working with TAF data would observe race/ethnicity information for. In contrast, Column 2 provides mortality rates using race/ethnicity information from Decennial/ACS data for beneficiaries who were missing race/ethnicity information in TAF data. This table indicates that those missing race/ethnicity in TAF but whose race/ethnicity we could find in Decennial/ACS data had higher mortality rates across all racial/ethnic groups. This provides some initial empirical evidence that those missing race/ethnicity were different from those who had this information in Medicaid data in terms of mortality outcomes.<sup>16</sup> These differences also persisted for the period including the pandemic (Columns 3 and 4). We also estimated average annual mortality rates for the detailed race/ethnicity variable (Table C-1).<sup>17</sup>

---

<sup>16</sup> Supplementary analyses showed that Medicaid beneficiaries missing race/ethnicity information in TAF data but with valid race/ethnicity information in Decennial/ACS data were, on average, older than Medicaid beneficiaries not missing race/ethnicity in TAF data. This is in part due to challenges assigning PIKs to children.

<sup>17</sup> We note that mortality rates presented throughout this paper were not adjusted for key sociodemographic characteristics. Our goal was to compare mortality rates within a given racial/ethnic group based on their missingness status in TAF rather than to compare mortality rates across racial/ethnic groups. Age distributions varied across racial/ethnic groups with Hispanic beneficiaries being, on average, younger relative to White beneficiaries. Due to this, the unadjusted mortality rate for Hispanic beneficiaries is much lower than the unadjusted mortality rate for White beneficiaries.

**Table 9. Average Annual Mortality Rate per 10,000 Beneficiaries by Missing Status and Time Period across Race/Ethnicity Groups**

	Before the Pandemic (January 2016 – February 2020)		Including the Pandemic (January 2016 - June 2021)		N	
	Not missing TAF race; TAF race	Missing TAF race; Decennial/ACS race	Not missing TAF race; TAF race	Missing TAF race; Decennial/ACS race	Not missing TAF race; TAF race	Missing TAF race; Decennial/ACS race
	(1)	(2)	(3)	(4)	(5)	(6)
<b>Hispanic</b>	42.08	49.80	46.13	55.77	21,120,000	2,901,000
<b>White</b>	131.60	160.40	130.20	157.90	34,060,000	5,466,000
<b>Black</b>	84.92	92.49	89.24	99.77	16,190,000	2,435,000
<b>Other</b>	68.16	77.11	73.61	80.90	5,929,000	1,063,000

Source: T-MSIS Analytic Files (TAF) data (2016); Census Numident (Q2, 2021); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

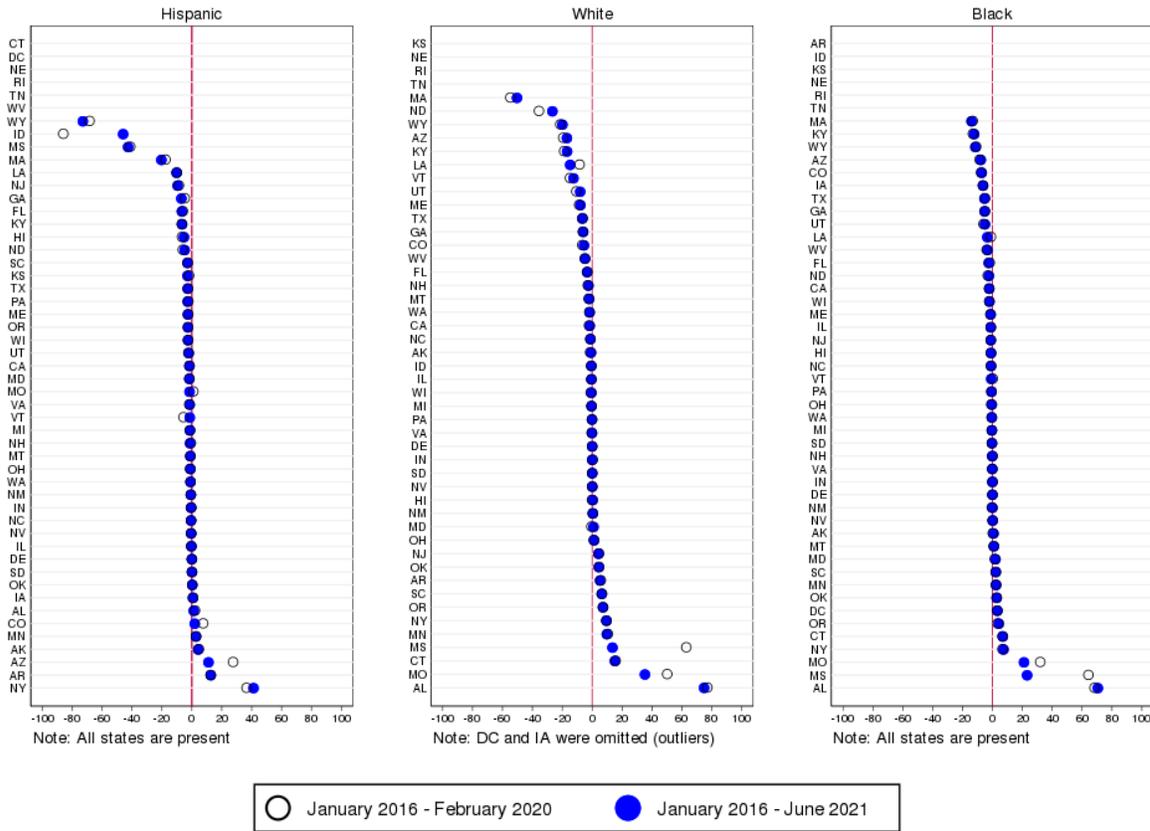
Notes: Mortality rates were not age adjusted. DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

Figure 6 and Figure 7 depict the differences in state-specific average annual mortality rates per 10,000 Medicaid beneficiaries for three racial/ethnic groups for TAF Race relative to Supplemental Race and Prioritized Race, respectively. Both figures show mortality rates for two periods – before the pandemic and including the pandemic – to assess whether state-specific mortality bias estimates varied between the periods. States with no information either did not have data for that race/ethnic category or had rates suppressed due to small denominators. We omitted several outlier states from these figures for easier viewing of the figures, as specified in the note for each figure.

Figure 6 shows the differences in the average annual mortality rates between TAF Race and Supplemental Race for each state and period. Negative values suggest an underestimation of mortality rates using TAF Race while positive values suggest an overestimation of mortality rates using TAF Race. The closer a difference is to zero the smaller the difference between TAF Race and Supplemental Race in estimating race-specific mortality rates. While mortality rates during both periods are presented in the graph, mortality rates provided in the text below cover the period including the pandemic (January 2016 – June 2021). For each racial/ethnic groups, most states hover close to zero, indicating very little difference in race-specific mortality between TAF Race and Supplemental Race. For these states, researchers focusing only on Medicaid beneficiaries with non-missing race/ethnicity information in TAF data are likely to produce unbiased estimates of race-specific mortality rates. However, there are some important exceptions. For example, for Hispanic ethnicity, Wyoming (-72.93), Idaho (-46), Mississippi (-42.66), and Massachusetts (-20.39) underestimated mortality rates using TAF Race relative to Supplemental Race. In contrast, for New York mortality rates were overestimated by 41.24 Hispanic beneficiaries per 10,000 when using TAF Race relative to Supplemental Race. For White beneficiaries, Iowa (-243.45, not pictured), Massachusetts (-50.3) and North Dakota (-26.6) underestimated mortality rates while D.C. (197.4, not pictured), Alabama (74.9) and Missouri (35.3) overestimated mortality rates using TAF Race relative to Supplemental Race. Finally, mortality rates for Black beneficiaries remained

comparable between the two approaches, with the exception of Alabama (70.5), Mississippi (23.2), and Missouri (21.1).

**Figure 6. Difference in Average Annual Race-Specific Mortality Rate per 10,000, TAF Race and Supplemental Race**



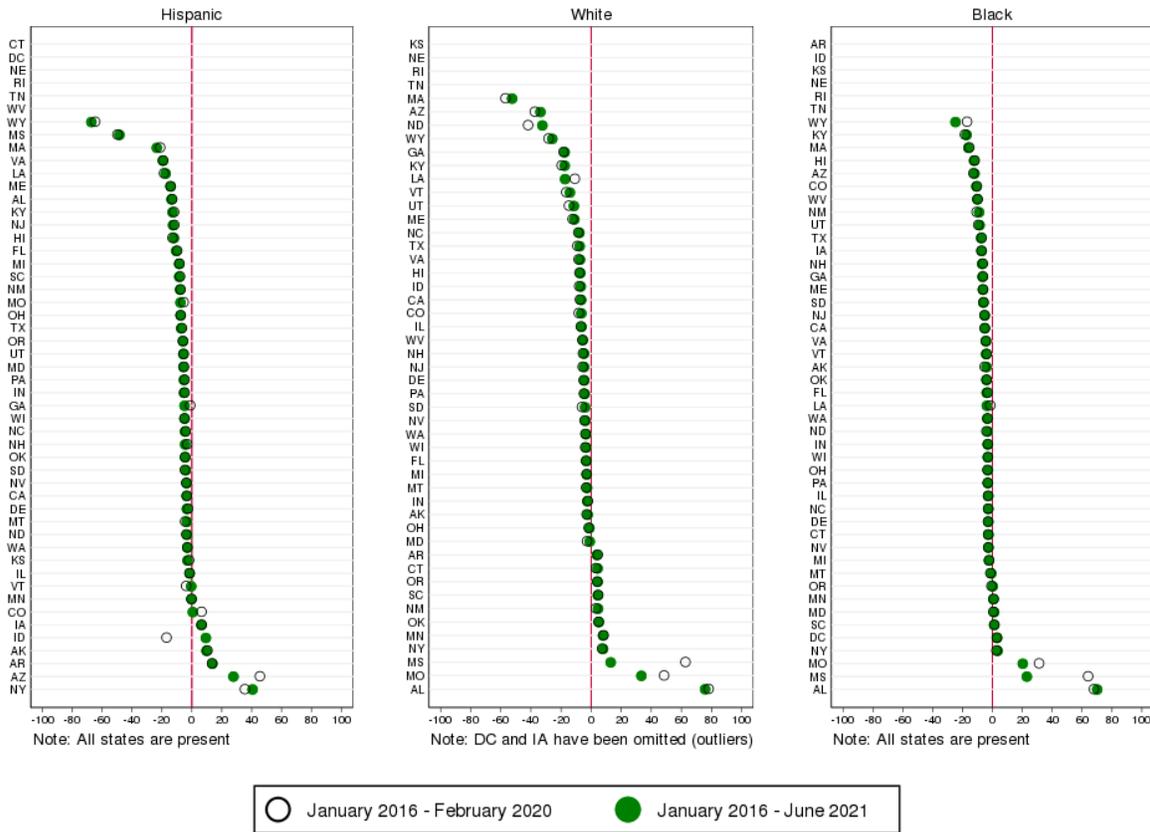
Source: T-MSIS Analytic Files (TAF) data (2016); Census Numident (Q2, 2021); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Note: Mortality rates were not age adjusted. Omitted values include DC (204.7 and 197.4) and IA (-262.65 and -243.45) for White beneficiaries (before the pandemic and including the pandemic, respectively). Differences are calculated as the annual mortality rate based on TAF Race minus the annual mortality rate based on the Supplemental Race. Positive values suggest that mortality rates were overestimated and negative values suggest that mortality rates were underestimated when relying on TAF Race. States with no information either did not have data for that race/ethnic category in TAF data or had rates suppressed due to small denominators. DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

The pandemic period did not impact the mortality rate bias estimation for most states. However, in a few instances, there were substantial differences not only between the different race/ethnicity variable approaches, but also by the period during which the mortality rates were calculated. For instance, Figure 6 shows that Mississippi had a substantial difference in White (63 and 13.6) and Black (64.2 and 23.2) mortality rate bias in the period before the pandemic relative to the period that included the pandemic, respectively. This suggests that the mortality rate for White and Black beneficiaries was overestimated in both periods when relying on TAF Race alone. But the magnitude of overestimation was exacerbated when focusing only on the period before the pandemic. Magnitude of the bias varied by state and race/ethnicity group and depended on the state-specific differences in the composition of beneficiaries with and without race/ethnicity information in TAF data.

Figure 7 shows the differences in the average annual mortality between TAF Race and Prioritized Race for each state. The closer a difference is to zero the smaller the difference between TAF Race and Prioritized Race in estimating race-specific mortality rates. Relative to Figure 6, there was more variation across states in the degree of difference in race-specific mortality rates. For Hispanic ethnicity, fewer states were hovering around zero and underestimation of Hispanic mortality using TAF Race relative to Prioritized Race occurred in most states. This was especially the case in Wyoming (-67.2), Mississippi (-48.46), and Massachusetts (-23.44). Only a couple of states substantially overestimated Hispanic mortality (i.e., Arizona and New York). For White beneficiaries, underestimation of mortality occurred in five states (i.e., Iowa, Massachusetts, Arizona, North Dakota, Wyoming), while overestimation of mortality occurred in Missouri (33.4), Alabama (75.8), and DC (204; not shown). Finally, for Black beneficiaries, there does not appear to be much difference in the mortality rate bias between the two approaches, with some exceptions, including Wyoming (-24.7), Mississippi (23), and Alabama (70.1).

**Figure 7. Difference in Average Annual Race-Specific Mortality Rate per 10,000, TAF Race and Prioritized Race**



Source: T-MSIS Analytic Files (TAF) data (2016); Census Numident (Q2, 2021); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Note: Mortality rates were not age adjusted. Omitted values include DC (210.7 and 204) and IA (-263.65 and -244.35) for White beneficiaries (before the pandemic and including the pandemic, respectively). Differences are calculated as the annual mortality rate based on TAF Race minus the annual mortality rate based on the Prioritized Race. Positive values suggest that mortality rates were overestimated and negative values suggest that mortality rates were underestimated when relying on TAF Race. States with no information either did not have data for that race/ethnic category in TAF data or had rates suppressed due to small denominators. DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

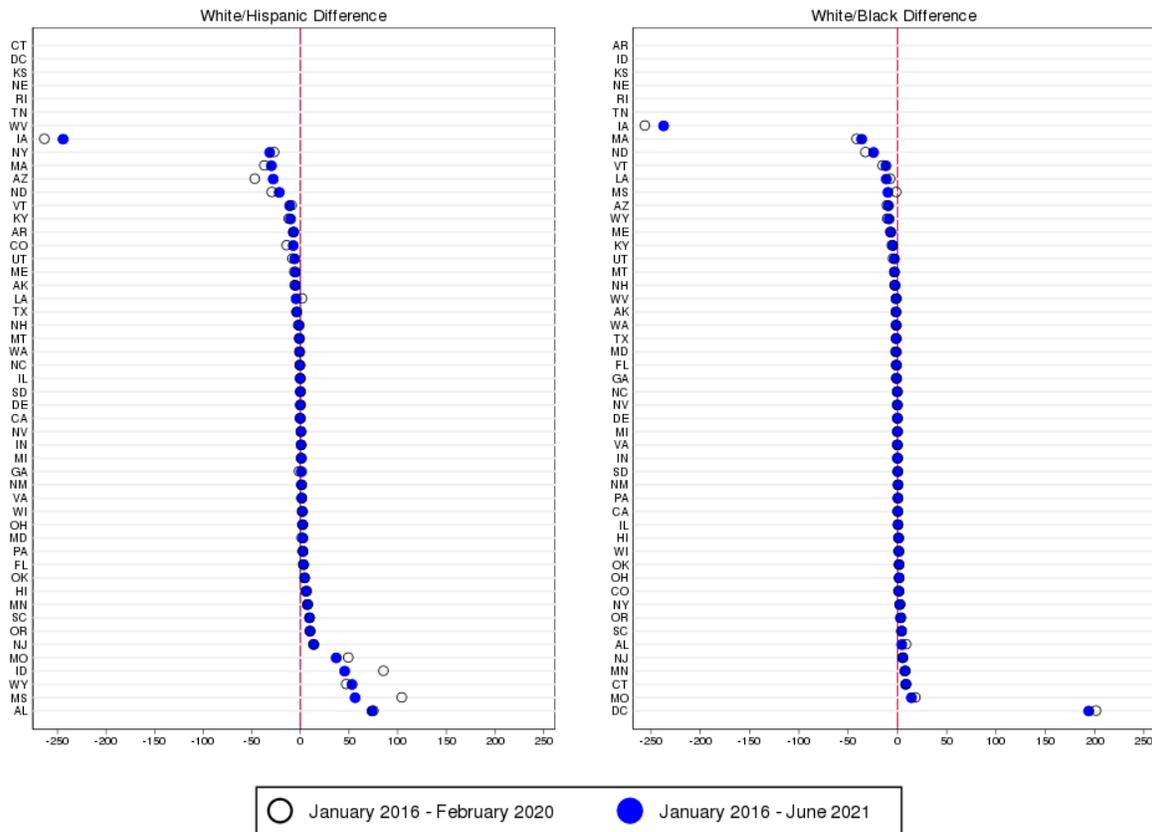
The majority of states did not appear to have high levels of bias being introduced into estimations of race-specific all-cause mortality. However, there were important differences based on whether we compared TAF Race to the Supplemental or Prioritized Race approach. Based on the prioritization assumptions behind the Prioritized Race approach, there was more movement in this approach, especially for Hispanic and White beneficiaries. For this analysis, we relied on crude rather than adjusted mortality rates. However, supplementary analyses using age adjusted mortality rates did not result in considerable change to the mortality rate bias observed. There were a few instances in which there was a large absolute change in mortality rate bias for a given state, but this was primarily restricted to states with high levels of missingness for a particular race/ethnicity group.

In addition to looking at race-specific mortality rates, we also investigated how bias could be introduced into the estimation of racial/ethnic disparities in mortality. To do this, we calculated the White/Hispanic and White/Black unadjusted mortality gaps. We did not age-adjust mortality rates and again caution the reader from making any conclusions about racial/ethnic mortality disparities based on the figures presented in this section. Our goal instead was to assess the change in disparities with and without the use of Decennial/ACS data for race/ethnicity information. Since we observed both over- and underestimation of mortality rates depending on state and racial/ethnic group, we had no prediction regarding the impact of the bias on mortality disparities (defined as differences in mortality rates between two race groups). If White and Hispanic mortality rates were underestimated to the same degree within a given state, then there would be no change in estimated mortality disparity for these two groups. On the other hand, there could be large differences in estimated disparities if the bias was sufficiently different between the two groups, especially if one group's mortality rate was overestimated while the other group's mortality rate was underestimated.

Figure 8 and Figure 9 compare the difference between White and Hispanic/Black mortality rates for both TAF Race relative to Supplemental Race and Prioritized Race, respectively. We calculated the racial/ethnic mortality disparity by subtracting the White and Hispanic/Black mortality difference for Supplemental Race (annual White mortality rate minus annual Hispanic/Black mortality rate) from the White and Hispanic/Black mortality difference for TAF Race (annual White mortality rate minus annual Hispanic/Black mortality rate). A value of zero indicated that there was no difference in the mortality disparity between race/ethnic groups across the two approaches (TAF Race and Supplemental Race). A positive value indicated that the estimated mortality disparity between racial/ethnic groups was being overestimated in TAF Race, while a negative value indicated that it was being underestimated in TAF Race. States with no information either did not have data for that racial/ethnic category or had rates suppressed due to small denominators.

In Figure 8, differences in the estimation of race/ethnic mortality disparities appear small, with a few important exceptions. For the White/Hispanic disparity, for example, in Iowa (-244.2), the racial/ethnic mortality disparity was substantially underestimated. The White/Hispanic mortality disparity was also being substantially overestimated in Missouri (36.82), Idaho (45.48), Wyoming (53.03), Mississippi (56.26), and Alabama (73.63). For the White/Black disparity, Iowa (-237.19), again, had underestimation of the disparity. Overestimation of the White/Black mortality rate disparity was also present in DC (194.1).

**Figure 8. Racial/Ethnic Disparities in Mortality, TAF Race and Supplemental Race**



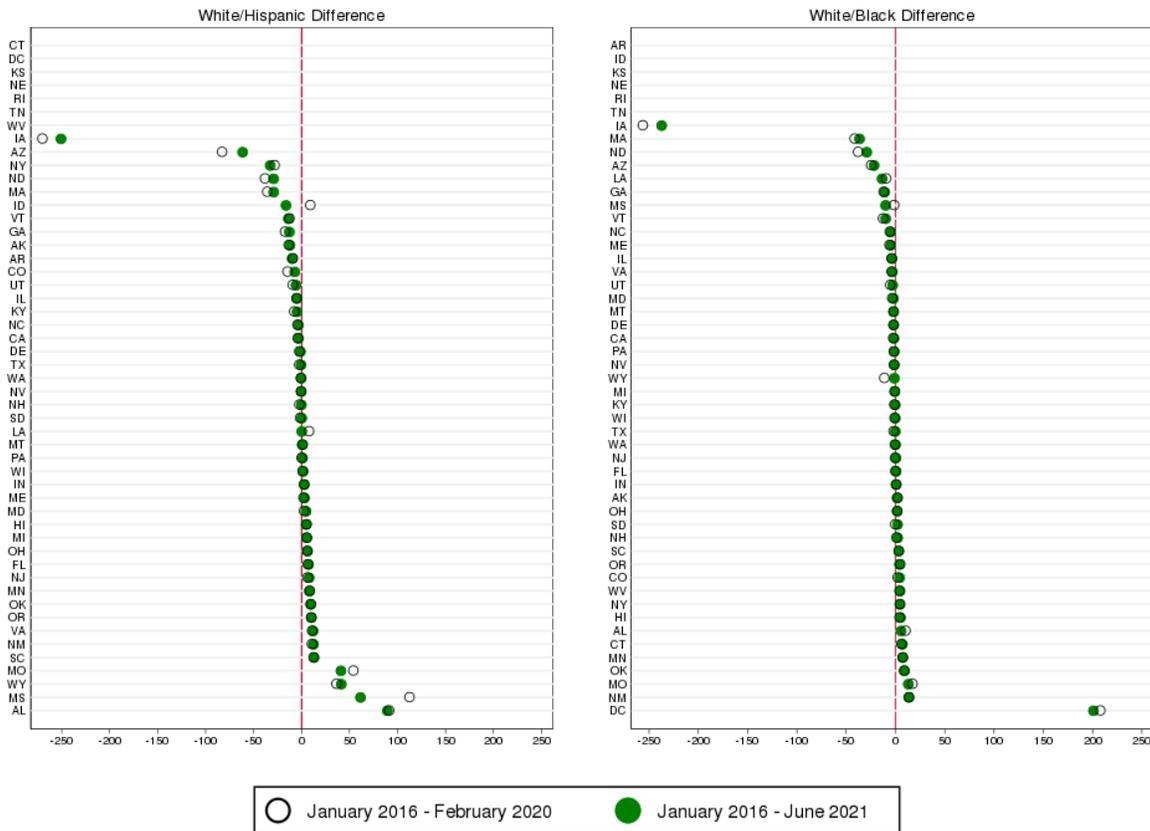
Source: T-MSIS Analytic Files (TAF) data (2016); Census Numident (Q2, 2021); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Note: Mortality rates were not age adjusted. The calculation for race/ethnic mortality disparity subtracted the White and Hispanic/Black mortality difference for Supplemental Race (annual mortality rate for White – annual mortality rate for Hispanic) from the White and Hispanic/Black mortality difference for TAF Race (annual mortality rate for White – annual mortality rate for Hispanic). Therefore, a value of zero indicates that there was no difference in the annual mortality rates between race/ethnic groups across the two approaches (TAF Race and Supplemental Race). A positive value indicates that the estimated mortality disparity between race/ethnic groups was being overestimated in TAF Race, while a negative value indicates that it was being underestimated in TAF Race. States with no information either did not have data for that race/ethnic category or had rates suppressed due to small denominators. DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

Relative to Figure 8, Figure 9 looks comparable in terms of the mortality disparity differences for White beneficiaries relative to Hispanic and Black beneficiaries with a few important exceptions. There was a higher degree of underestimation of the White/Hispanic mortality disparity in Arizona (-61.46). Additionally, Idaho (-16.22) flipped to an underestimation of the White/Hispanic disparity when Prioritized Race was used as the comparison. Beyond a small amount of reordering across states, there were essentially no differences in the estimation of White/Black mortality disparities between Supplemental Race and Prioritized Race. Overall, the results for both Figure 8 and Figure 9 suggest that using Medicaid data from most states does not seem to be producing

biased estimates of racial/ethnic disparities in mortality for either Hispanic or Black beneficiaries relative to White beneficiaries. There are some notable exceptions with Iowa and DC being extreme outliers using both approaches. Overall, these findings suggest that researchers should be attentive to how this mortality rate bias may be uniquely transferred to the estimation of racial/ethnic disparities when comparing race/ethnicity groups.

**Figure 9. Racial/Ethnic Disparities in Mortality, TAF Race and Prioritized Race**



Source: T-MSIS Analytic Files (TAF) data (2016); Census Numident (Q2, 2021); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Note: Mortality rates were not age adjusted. The calculation for race/ethnic mortality disparity subtracted the White and Hispanic/Black mortality difference for Prioritized Race (annual mortality rate for White – annual mortality rate for Hispanic) from the White and Hispanic/Black mortality difference for TAF Race (annual mortality rate for White – annual mortality rate for Hispanic). Therefore, a value of zero indicates that there was no difference in the annual mortality rates between race/ethnic groups across the two approaches (TAF Race and Prioritized Race). A positive value indicates that the estimated mortality disparity between race/ethnic groups was being overestimated in TAF Race, while a negative value indicates that it was being underestimated in TAF Race. States with no information either did not have data for that race/ethnic category or had rates suppressed due to small denominators. DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

Beyond estimating differences across race/ethnicity groups and states, mortality rates also varied considerably by age. Medicaid beneficiaries were, on average, younger and, thus, generally had lower mortality rates compared to the general population. We conducted supplementary analyses

that age-stratified our main set of analyses (Figure C-3 through Figure C-8). In both approaches, there was considerably more bias in the estimation of mortality rates for older Medicaid beneficiaries relative to younger ones.

Overall, our results suggest that a variety of factors including race/ethnicity, age, and state must be considered when thinking about bias in estimates of all-cause mortality among Medicaid beneficiaries. Most importantly, there were differences across states in terms of mortality rate bias and these differences varied by the race/ethnicity group under investigation. For Hispanic ethnicity, Arizona, Idaho, Massachusetts, Mississippi, New York, and Wyoming exhibited exceptionally high rates of mortality rate bias. For White race, Alabama, Arizona, Iowa, DC, Massachusetts, Missouri, North Dakota, and Wyoming exhibited exceptionally high rates of mortality rate bias. Finally, for Black race, Alabama, Mississippi, Missouri, and Wyoming demonstrated exceptionally high rates of mortality rate bias.

## **Conclusions**

It is increasingly important to systematically collect high-quality race and ethnicity data to further bolster research focused on racial/ethnic disparities in health and health care (Executive Office of the President 2021; James et al. 2021; National Research Council 2004; Institute of Medicine 2009). This is especially true for Medicaid data (Brown-Podgorski, Roberts, and Schpero 2022; MACPAC 2021b) since there is considerable variation in the completeness of race/ethnicity data for beneficiaries across states (Khan et al. 2019). This research assessed the utility of linking Medicaid enrollment data and Census Bureau microdata sources (i.e., Decennial Census and ACS) as one way to enhance race/ethnicity information in Medicaid data. Additionally, we evaluated the degree to which using data with missing race/ethnicity information results in biased estimates of all-cause mortality. Our hope is that this paper can serve as a detailed state-level reference guide for researchers interested in utilizing race/ethnicity information in Medicaid data.

Our research had several key findings. First, similar to previous research, linking Medicaid enrollment data and Decennial/ACS data is not only feasible, but also effective in providing race/ethnicity information for beneficiaries missing this information in Medicaid data (Fernandez et al. 2015). It was possible to fill in race/ethnicity information for 63.62% of beneficiaries who were missing this information nationally, resulting in only 7.11% of all beneficiaries with unknown race/ethnicity after the linkage to Decennial/ACS data (down from 19.47%). Our success in using Decennial/ACS data to fill in missing race/ethnicity varied considerably across states. Second, in terms of concordance, we found that 87.54% of beneficiaries with race/ethnicity information available in both sources had the same race/ethnicity listed, falling in line with previous research (Fernandez et al. 2015). Here too, concordance rates varied considerably across states.

Finally, in terms of the bias assessment, our results demonstrated that in most instances, missing race/ethnicity in Medicaid data did not appear to pose a substantial problem for the estimation of race-specific mortality rates. However, there were important exceptions based on race/ethnicity, state, and age. In terms of race/ethnicity, White and Hispanic beneficiaries saw substantial misestimations of mortality rates while the estimation of mortality rates for Black beneficiaries remained relatively consistent across the two modified race/ethnicity approaches. In terms of state

analyses, most states did not exhibit high levels of mortality rate bias, but there were some exceptions. We also found that older beneficiaries exhibited higher rates of mortality rate bias than younger beneficiaries. Therefore, race/ethnicity, state, and age are all factors that must be considered in tandem when thinking about how best to proceed with health research focused on racial/ethnic disparities that rely on Medicaid enrollment data.

Medicaid is not the only data that have incomplete race/ethnicity information and other health records face similar issues (Branham et al. 2022; Klinger et al. 2015; Krieger et al. 2020; Polubriaginof et al. 2019). Recent research has attempted to solve the problem of missing race/ethnicity information in health records data including by using imputation techniques (Grundmeier et al. 2015; Krieger et al. 2008; Labgold et al. 2021; Sorbero et al. 2022), natural language processing algorithms (Sholle et al. 2019), and data linkage strategies (Fernandez et al. 2015; Kressin et al. 2003; McAlpine et al. 2007). Our study contributes to this growing area of research focused on improving the quality and usability of health records data.

The methods employed and findings uncovered in this research have considerable potential beyond Medicaid data. For example, electronic health records (EHRs), health registries, inpatient/outpatient data, and emergency department data can similarly be linked to Census Bureau microdata sources. Some recent work has already successfully linked a variety of health records with Census Bureau microdata, including emergency department visit data from the Utah Department of Health (Powers et al. 2021) and patient-level EHRs from a healthcare system in the Southeast (Udalova et al. 2022) linked to ACS data. The potential of linking health records to restricted Census Bureau microdata is extensive given the unique resources of federal microdata. This potential is only further enhanced given the continuous and compulsory nature of health records especially in light of increasing challenges with regards to collecting new survey data (Conrad, Keusch, and Schober 2021; Robert M. Groves 2011; Mooney and Pejaver 2018).

Health records data also have the potential to advance the mission of the Census Bureau. Hard-to-reach populations may be captured by health data, even if they are not captured by any other survey, census, or administrative records data. For instance, everyone needs healthcare at some point in their lives and emergency departments are required to treat everyone regardless of their ability to pay for care or their immigration status. Health records can also be used to validate similar concepts captured in Census Bureau data, including demographic characteristics. With a plethora of demographic, social, contextual, and economic data available at the Census Bureau, linkages with health records can elucidate the role of social determinants of health and the pathways that lead to disparities in health.

This study focused on mortality outcomes as a case study since mortality is consistently measured, readily available, and is one of the most studied health outcomes in health disparities research. However, mortality is not the only health outcome that may be biased when using data with missing race/ethnicity information. The findings presented here may not be generalizable to other health outcomes. Investigation of additional health outcomes is a promising arena for future research. Additionally, although we focused exclusively on race/ethnicity here, there are other demographic elements that are inconsistently captured in Medicaid data (Khan et al. 2019).

This study did have some limitations. After linkage, 7.11% of Medicaid beneficiaries still did not have any race/ethnicity information. This is likely a unique group in terms of both demographic characteristics and health outcomes. Previous research has discussed the implications and potential limitations of incomplete data linkages (Bohensky et al. 2010). However, given the differences between individuals missing on race/ethnicity, relative to those who are not (Ennis et al. 2018; Fernandez et al. 2015), we suspect that this group of beneficiaries may have their own set of unique characteristics. For example, they might be less likely to have formal ties to the labor market or be recipients of public programs. They are likely the individuals who are hard to reach during surveys and decennial censuses. This conjecture is based on the way the Census Bureau assigns the linkage keys. We suspect that this group may even have worse health outcomes than the group of beneficiaries we could find in Decennial/ACS data. Therefore, we suspect that our overall findings of mortality rate bias may be conservative.

Our findings demonstrate significant potential for using Census Bureau microdata to complement existing health records that commonly lack important demographic characteristics, such as race/ethnicity. We find that enhancing race/ethnicity information in Medicaid enrollment data with population-level Census Bureau microdata is feasible and can advance our understanding of population health. Medicaid data and health records, more generally, can advance the goals of the Census Bureau by providing an opportunity to further leverage already available resources by exploring important population health topics. This research has shown how unique data linkages can be used strategically to advance the field of population health.

## References

- Baumgartner, Jesse, Sara Collins, David Radley, and Susan Hayes. 2020. “How the Affordable Care Act (ACA) Has Narrowed Racial and Ethnic Disparities in Insurance Coverage and Access to Health Care, 2013-18.” *Health Services Research* 55 (S1): 56–57. <https://doi.org/10.1111/1475-6773.13406>.
- Bilaver, Lucy A., Sarah A. Sobotka, and David S. Mandell. 2021. “Understanding Racial and Ethnic Disparities in Autism-Related Service Use Among Medicaid-Enrolled Children.” *Journal of Autism and Developmental Disorders* 51 (9): 3341–55. <https://doi.org/10.1007/s10803-020-04797-6>.
- Bohensky, Megan A., Damien Jolley, Vijaya Sundararajan, Sue Evans, David V. Pilcher, Ian Scott, and Caroline A. Brand. 2010. “Data Linkage: A Powerful Research Tool with Potential Problems.” *BMC Health Services Research* 10 (1): 346. <https://doi.org/10.1186/1472-6963-10-346>.
- Bond, Brittany, J. David Brown, Adela Luque, and Amy O’Hara. 2014. “The Nature of the Bias When Studying Only Linkable Person Records: Evidence from the American Community Survey.” Center for Administrative Records Research and Applications Working Paper CARRA-WP-2014-08. U.S. Census Bureau.
- Branham, D. Keith, Kenneth Finegold, Lucy Chen, Melony Sorbero, Roald Euller, Marc N. Elliott, and Benjamin D. Sommers. 2022. “Trends in Missing Race and Ethnicity Information After Imputation in HealthCare.Gov Marketplace Enrollment Data, 2015-2021.” *JAMA Network Open* 5 (6): e2216715. <https://doi.org/10.1001/jamanetworkopen.2022.16715>.
- Brown-Podgorski, Brittany, Eric Roberts, and William Schpero. 2022. “Improving Medicaid Data To Advance Racial And Ethnic Health Equity In The United States | Health Affairs Forefront.” Health Affairs. 2022. <https://www.healthaffairs.org/doi/10.1377/forefront.20220517.465853/full/>.
- Buescher, Paul A., J. Timothy Whitmire, and Barbara Pullen-Smith. 2010. “Medical Care Costs for Diabetes Associated with Health Disparities Among Adult Medicaid Enrollees in North Carolina.” *North Carolina Medical Journal* 71 (4): 319–24. <https://doi.org/10.18043/ncm.71.4.319>.
- Cantor, Joel C., Sujoy Chakravarty, Jose Nova, Taiisa Kelly, Derek Delia, Emmy Tiderington, and Richard W. Brown. 2020. “Medicaid Utilization and Spending among Homeless Adults in New Jersey: Implications for Medicaid-Funded Tenancy Support Services.” *The Milbank Quarterly* 98 (1): 106–30. <https://doi.org/10.1111/1468-0009.12446>.
- Centers for Medicare & Medicaid Services. 2021. “Chronic Conditions Warehouse User Guide.” Chronic Conditions Data Warehouse. November 2021. <https://www2.ccwdata.org/web/guest/user-documentation>.
- . 2022. “DQ ATLAS.” 2022. <https://www.medicaid.gov/dq-atlas>.

- Compton, Elizabeth, Michael Bentley, Sharon Ennis, and Sonya Rastogi. 2013. “2010 Census Race and Hispanic Origin Alternative Questionnaire Experiment.” 2010 Census Planning Memoranda Series CPEX-211. Washington, D.C.: U.S. Census Bureau.
- Conrad, Frederick G., Florian Keusch, and Michael F. Schober. 2021. “New Data in Social and Behavioral Research.” *Public Opinion Quarterly* 85 (S1): 253–63. <https://doi.org/10.1093/poq/nfab027>.
- Davis, Melinda M., Stephanie Renfro, Robyn Pham, Kristen Hassmiller Lich, Jackilen Shannon, Gloria D. Coronado, and Stephanie B. Wheeler. 2017. “Geographic and Population-Level Disparities in Colorectal Cancer Testing: A Multilevel Analysis of Medicaid and Commercial Claims Data.” *Preventive Medicine* 101: 44–52. <https://doi.org/10.1016/j.ypmed.2017.05.001>.
- Ennis, Sharon R., Sonya R. Porter, James M. Noon, and Ellen Zapata. 2018. “When Race and Hispanic Origin Reporting Are Discrepant across Administrative Records and Third Party Sources: Exploring Methods to Assign Responses.” *Statistical Journal of the IAOS* 34: 179–89. <https://doi.org/DOI 10.3233/SJI-170374>.
- Executive Office of the President. 2021. “Executive Order on Advancing Racial Equity and Support for Underserved Communities Through the Federal Government.” Washington, D.C.: The White House. <https://www.federalregister.gov/documents/2021/01/25/2021-01753/advancing-racial-equity-and-support-for-underserved-communities-through-the-federal-government>.
- Fabius, Chanee D., Kali S. Thomas, Tingting Zhang, Jessica Ogarek, and Theresa I. Shireman. 2018. “Racial Disparities in Medicaid Home and Community-Based Service Utilization and Expenditures among Persons with Multiple Sclerosis.” *BMC Health Services Research* 18 (773). <https://doi.org/10.1186/s12913-018-3584-x>.
- Fernandez, Leticia E., Sonya Rastogi, Sharon R. Ennis, and James M. Noon. 2015. “Evaluating Race and Hispanic Origin Responses of Medicaid Participants Using Census Data.” Center for Administrative Records Research and Applications Working Paper CARRA-WP-2015-01. Washington, D.C.: U.S. Census Bureau.
- Finlay, Keith, and Katie R. Genadek. 2021. “Measuring All-Cause Mortality with the Census Numident File.” *American Journal of Public Health* 111 (S2): S141–48. <https://doi.org/10.2105/AJPH.2021.306217>.
- Groves, Robert M. 2011. “Three Eras of Survey Research.” *Public Opinion Quarterly* 75 (5): 861–71. <https://doi.org/10.1093/poq/nfr057>.
- Grundmeier, Robert W, Lihai Song, Mark J Ramos, Alexander G Fiks, Marc N Elliott, Allen Fremont, Wilson Pace, Richard C Wasserman, and Russell Localio. 2015. “Imputing Missing Race/Ethnicity in Pediatric Electronic Health Records: Reducing Bias with Use of U.S. Census Location and Surname Data.” *Health Services Research* 50 (4): 946–60. <https://doi.org/10.1111/1475-6773.12295>.

- Heflinger, Craig Anne, Jamie Chatman, and Robert C. Saunders. 2006. “Racial and Gender Differences in Utilization of Medicaid Substance Abuse Services Among Adolescents.” *Psychiatric Services* 57 (4): 504–11. <https://doi.org/10.1176/ps.2006.57.4.504>.
- Horvitz-Lennon, Marcela, Rita Volya, Rachel Garfield, Julie M. Donohue, Judith R. Lave, and Sharon-Lise T. Normand. 2015. “Where You Live Matters: Quality and Racial/Ethnic Disparities in Schizophrenia Care in Four State Medicaid Programs.” *Health Services Research* 50 (5): 1710–29. <https://doi.org/10.1111/1475-6773.12296>.
- Institute of Medicine. 2009. “Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement.” Washington, D.C.: The National Academic Press.
- James, Cara, Smita Pamar, Sarah Hudson Scholle, Philip Saynisch, Jeni Soucie, and Barbara Lyons. 2021. “Federal Action Is Needed to Improve Race and Ethnicity Data in Health Programs.” Washington, D.C.: Grantmakers in Health and National Committee for Quality Assurance. <https://www.gih.org/wp-content/uploads/2021/10/GIH-Commonwealth-Fund-federal-data-report-part-1.pdf>.
- Khan, Tsnuva, Esa Eslami, Lello Guluma, Kimberly Proctor, and Jessie Parker. 2019. “Quantifying the Distribution and Completeness of Select Demographic Variables in 2016.” TAF DQ Brief #4121. Baltimore, MD: CMS.
- Kirby, James B., and Toshiko Kaneda. 2010. “Unhealthy and Uninsured: Exploring Racial Differences in Health and Health Insurance Coverage Using a Life Table Approach.” *Demography* 47 (4): 1035–51. <https://doi.org/10.1007/BF03213738>.
- Klinger, Elissa V., Sara V. Carlini, Irina Gonzalez, Stella St. Hubert, Jeffrey A. Linder, Nancy A. Rigotti, Emily Z. Kontos, Elyse R. Park, Lucas X. Marinacci, and Jennifer S. Haas. 2015. “Accuracy of Race, Ethnicity, and Language Preference in an Electronic Health Record.” *Journal of General Internal Medicine* 30 (6): 719–23. <https://doi.org/10.1007/s11606-014-3102-8>.
- Kressin, Nancy R., Bei-Hung Chang, Ann Hendricks, and Lewis E. Kazis. 2003. “Agreement Between Administrative Data and Patients’ Self-Reports of Race/Ethnicity.” *American Journal of Public Health* 93 (10): 1734–39. <https://doi.org/10.2105/AJPH.93.10.1734>.
- Krieger, Nancy, Jarvis T. Chen, James H. Ware, and Afamia Kaddour. 2008. “Race/Ethnicity and Breast Cancer Estrogen Receptor Status: Impact of Class, Missing Data, and Modeling Assumptions.” *Cancer Causes & Control* 19 (10): 1305–18. <https://doi.org/10.1007/s10552-008-9202-1>.
- Krieger, Nancy, Gregg Gonsalves, Mary T. Bassett, William Hanage, and Harlan M. Krumholz. 2020. “The Fierce Urgency of Now: Closing Glaring Gaps In US Surveillance Data On COVID-19.” *Health Affairs*. 2020. <https://www.healthaffairs.org/doi/10.1377/forefront.20200414.238084/full/>.
- Labgold, Katie, Sarah Hamid, Sarita Shah, Neel R. Gandhi, Allison Chamberlain, Fazle Khan, Shamimul Khan, et al. 2021. “Estimating the Unknown: Greater Racial and Ethnic

- Disparities in COVID-19 Burden After Accounting for Missing Race and Ethnicity Data.” *Epidemiology* 32 (2): 157–61. <https://doi.org/10.1097/EDE.0000000000001314>.
- Lee, Hyunjung, and Frank W. Porell. 2020. “The Effect of the Affordable Care Act Medicaid Expansion on Disparities in Access to Care and Health Status.” *Medical Care Research and Review* 77 (5): 461–73. <https://doi.org/10.1177/1077558718808709>.
- Liebler, Carolyn A., Sonya R. Porter, Leticia E. Fernandez, James M. Noon, and Sharon R. Ennis. 2017. “America’s Churning Races: Race and Ethnicity Response Changes Between Census 2000 and the 2010 Census.” *Demography* 54 (1): 259–84. <https://doi.org/10.1007/s13524-016-0544-0>.
- Lillie-Blanton, Marsha, and Catherine Hoffman. 2005. “The Role Of Health Insurance Coverage In Reducing Racial/Ethnic Disparities In Health Care.” *Health Affairs* 24 (2): 398–408. <https://doi.org/10.1377/hlthaff.24.2.398>.
- MACPAC. 2021a. “MACStats Medicaid and CHIP Data Book December 2021.” Washington, DC: Medicaid and CHIP Payment and Access Commission. <https://www.macpac.gov/wp-content/uploads/2021/12/MACStats-Medicaid-and-CHIP-Data-Book-December-2021.pdf>.
- . 2021b. “Re: Request for Information: Methods and Leading Practices for Advancing Equity and Support for Underserved Communities through Government Action, Federal Register, Vol 86, No. 85, May 5, 2021,” 2021. <https://www.macpac.gov/wp-content/uploads/2021/07/Comment-on-RFI-on-Advancing-Equity.pdf>.
- Mahajan, Shiwani, César Caraballo, Yuan Lu, Javier Valero-Elizondo, Daisy Massey, Amarnath R. Annapureddy, Brita Roy, et al. 2021. “Trends in Differences in Health Status and Health Care Access and Affordability by Race and Ethnicity in the United States, 1999-2018.” *JAMA* 326 (7): 637–48. <https://doi.org/10.1001/jama.2021.9907>.
- Marton, James, Aaron Yelowitz, Meredith Shores, and Jeffery C. Talbert. 2016. “Does Medicaid Managed Care Help Equalize Racial and Ethnic Disparities in Utilization?” *Health Services Research* 51 (3): 872–91. <https://doi.org/10.1111/1475-6773.12396>.
- McAlpine, Donna D., Timothy J. Beebe, Michael Davern, and Kathleen T. Call. 2007. “Agreement between Self-Reported and Administrative Race and Ethnicity Data among Medicaid Enrollees in Minnesota: Self-Reported and Administrative Race and Ethnicity Data.” *Health Services Research* 42 (6p2): 2373–88. <https://doi.org/10.1111/j.1475-6773.2007.00771.x>.
- McConnell, K. John, Christina J. Charlesworth, Thomas H. A. Meath, Rani M. George, and Hyunjee Kim. 2018. “Oregon’s Emphasis on Equity Shows Signs of Early Success For Black And American Indian Medicaid Enrollees.” *Health Affairs* 37 (3): 386–93. <https://doi.org/10.1377/hlthaff.2017.1282>.

- Melendez, Mitzi, Suzanne Campanella, Samantha Rosner, Tabitha Pyatt, Bryan Gustafson, Jordan Katz, and Emily Tcheng. 2022. “The State of the Collection of Race, Ethnicity, and Language Data in Medicaid.” NORC at the University of Chicago.
- Miyawaki, Michael Hajime. 2016. “Part-Latinos and Racial Reporting in the Census: An Issue of Question Format?” *Sociology of Race and Ethnicity* 2 (3): 289–306. <https://doi.org/10.1177/2332649215613531>.
- Mooney, Stephen J., and Vikas Pejaver. 2018. “Big Data in Public Health: Terminology, Machine Learning, and Privacy.” *Annual Review of Public Health* 39 (1): 95–112. <https://doi.org/10.1146/annurev-publhealth-040617-014208>.
- National Research Council. 2004. *Eliminating Health Disparities: Measurement and Data Needs*. Edited by Panel on DHHS Collection of Race and Ethnicity Data, Michele Ver Ploeg, and Edward Perrin. Washington, D.C.: The National Academies Press. <https://doi.org/10.17226/10979>.
- Ng, Judy H., Faye Ye, Lauren M. Ward, Samuel C. “Chris” Haffer, and Sarah Hudson Scholle. 2017. “Data On Race, Ethnicity, And Language Largely Incomplete For Managed Care Plan Members.” *Health Affairs* 36 (3): 548–52. <https://doi.org/10.1377/hlthaff.2016.1044>.
- Office of Management and Budget. 1997. “Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity.” The White House. 1997. <https://obamawhitehouse.archives.gov/node/15626>.
- Polubriaginof, Fernanda C G, Patrick Ryan, Hojjat Salmasian, Andrea Wells Shapiro, Adler Perotte, Monika M Safford, George Hripcsak, Shaun Smith, Nicholas P Tatonetti, and David K Vawdrey. 2019. “Challenges with Quality of Race and Ethnicity Data in Observational Databases.” *Journal of the American Medical Informatics Association* 26 (8–9): 730–36. <https://doi.org/10.1093/jamia/ocz113>.
- Polyakova, Maria, Victoria Udalova, Geoffrey Kocks, Katie Genadek, Keith Finlay, and Amy N. Finkelstein. 2021. “Racial Disparities in Excess All-Cause Mortality During the Early COVID-19 Pandemic Varied Substantially Across States.” *Health Affairs (Project Hope)* 40 (2): 307–16. <https://doi.org/10.1377/hlthaff.2020.02142>.
- Powers, David, Sara Robinson, Edward Berchick, J. Alex Branham, Lucinda Dalzell, Lorelle Dennis, Kristi Eckerson, et al. 2021. “Evaluating the Utility of Emergency Department Encounter Data and Examining Social Determinants of Emergency Department Utilization in Utah.” SEHSD Working Paper Series 2021–07. Washington, D.C.: U.S. Census Bureau.
- Rastogi, Sonya, and Amy O’Hara. 2012. “2010 Census Match Study.” 2010 Planning Memoranda Series No. 247. Washington, D.C.: U.S. Census Bureau.
- Ringwalt, Chris, Andrew W. Roberts, Hallam Gugelmann, and Asheley Cockrell Skinner. 2015. “Racial Disparities Across Provider Specialties in Opioid Prescriptions Dispensed to

- Medicaid Beneficiaries with Chronic Noncancer Pain.” *Pain Medicine* 16 (4): 633–40. <https://doi.org/10.1111/pme.12555>.
- Samnaliev, Mihail, Mark P. McGovern, and Robin E. Clark. 2009. “Racial/Ethnic Disparities in Mental Health Treatment in Six Medicaid Programs.” *Journal of Health Care for the Poor and Underserved* 20 (1): 165–76. <https://doi.org/10.1353/hpu.0.0125>.
- Saperstein, Aliya, and Andrew M. Penner. 2012. “Racial Fluidity and Inequality in the United States.” *American Journal of Sociology* 118 (3): 676–727. <https://doi.org/10.1086/667722>.
- Sholle, Evan T, Laura C Pinheiro, Prakash Adekkanattu, Marcos A Davila III, Stephen B Johnson, Jyotishman Pathak, Sanjai Sinha, et al. 2019. “Underserved Populations with Missing Race Ethnicity Data Differ Significantly from Those with Structured Race/Ethnicity Documentation.” *Journal of the American Medical Informatics Association* 26 (8–9): 722–29. <https://doi.org/10.1093/jamia/ocz040>.
- Silber, Jeffrey H., Paul R. Rosenbaum, Shawna R. Calhoun, Joseph G. Reiter, Alexander S. Hill, James P. Guevara, Joseph J. Zorc, and Orit Even-Shoshan. 2017. “Racial Disparities in Medicaid Asthma Hospitalizations.” *Pediatrics* 139 (1): e20161221. <https://doi.org/10.1542/peds.2016-1221>.
- Smedley, Brian D., Adrienne Y. Stith, and Alan R. Nelson, eds. 2003. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. Washington, D.C.: National Academies Press. <https://doi.org/10.17226/12875>.
- Sorbero, Melony E, Roald Euller, Aaron Kofner, and Marc N Elliott. 2022. “Imputation of Race and Ethnicity in Health Insurance Marketplace Enrollment Data, 2015–2022 Open Enrollment Periods.” RR-A1853-1. Santa Monica, CA: RAND Corporation. [https://www.rand.org/pubs/research\\_reports/RRA1853-1.html](https://www.rand.org/pubs/research_reports/RRA1853-1.html).
- Staras, Stephanie A.S., Susan T. Vadaparampil, Laura T. Haderxhanaj, and Elizabeth A. Shenkman. 2010. “Disparities in Human Papillomavirus Vaccine Series Initiation Among Adolescent Girls Enrolled in Florida Medicaid Programs, 2006–2008.” *The Journal of Adolescent Health* 47 (4): 381–88. <https://doi.org/10.1016/j.jadohealth.2010.07.028>.
- Tangka, Florence K., Sujha Subramanian, Lee Rivers Mobley, Sonja Hoover, Jiantong Wang, Ingrid J. Hall, and Simple D. Singh. 2017. “Racial and Ethnic Disparities among State Medicaid Programs for Breast Cancer Screening.” *Preventive Medicine* 102: 59–64. <https://doi.org/10.1016/j.ypmed.2017.06.024>.
- Terry, Rodney L., and Marissa Fond. 2013. “Experimental U.S. Census Bureau Race and Hispanic Origin Survey Questions: Reactions from Spanish Speakers.” *Hispanic Journal of Behavioral Sciences* 35 (4): 524–41. <https://doi.org/10.1177/0739986313499007>.
- Udalova, Victoria, Timothy S. Carey, Paul Roman Chelminski, Lucinda Dalzell, Patricia Knoepp, Joanna Motro, and Barbara Entwisle. 2022. “Linking Electronic Health Records

- to the American Community Survey: Feasibility and Process.” *American Journal of Public Health*, April, e1–8. <https://doi.org/10.2105/AJPH.2022.306783>.
- U.S. Department of Health and Human Services. 2020. “Social Determinants of Health.” Healthy People 2030. <https://health.gov/healthypeople/priority-areas/social-determinants-health>.
- Wagner, Deborah, and Mary Layne. 2014. “The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications’ (CARRA) Record Linkage Software.” CARRA Working Paper Series. U.S. Census Bureau.
- Whitney, Melissa, Ellen Singer, Allison Barrett, Kimberly Proctor, and Jessie Parker. 2020. “Unique Beneficiary Identifiers in the TAF RIF in 2016.” TAF DQ Brief #3031. Baltimore, MD: CMS.
- Zhang, Shun, Kathryn Cardarelli, Ruth Shim, Jiali Ye, Karla L. Booker, and George Rust. 2013. “Racial Disparities in Economic and Clinical Outcomes of Pregnancy Among Medicaid Recipients.” *Maternal and Child Health Journal* 17 (8): 1518–25. <https://doi.org/10.1007/s10995-012-1162-0>.
- Zylla, Emily, and Elizabeth Lukanen. 2021. “Collection of Race, Ethnicity, Language (REL) Data in Medicaid Applications: A 50-State Review of the Current Landscape.” State Health Access Data Assistance at the University of Minnesota.

## Supplemental Materials

### Appendix A. Data Linkage Supplement

The Data Linkage Supplement provides supplementary materials for the Data Linkage portion of the analysis. Table A-1 is similar to Table 4 but provides the race/ethnicity distribution of beneficiaries for the detailed race/ethnicity variable. The other race category is now broken out into Asian, American Indian/Alaskan Native (AIAN), Native Hawaiian/Pacific Islander (NHPI), and beneficiaries identifying as two or more races. Among these newly observed race/ethnicity groups, Asian beneficiaries had the lowest PIK rate (98.32%). AIAN beneficiaries had the highest linkage rate (73.96%) and those identifying as two or more races had the lowest linkage rate (59.51%).

**Table A-1. Detailed Race/Ethnicity Distribution Based on 2016 TAF Data**

	All Observations		Observations with PIKs		Observations Linked to Decennial/ACS		
	N	Column %	N	PIK Rate (%)	N	Linkage Rate (%)	Linkage Rate (%), Conditional on PIK
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<b>Hispanic</b>	21,120,000	22.00	19,730,000	93.42	13,030,000	61.70	66.04
<b>White</b>	34,060,000	35.48	33,900,000	99.53	26,880,000	78.92	79.29
<b>Black</b>	16,190,000	16.87	16,090,000	99.38	11,380,000	70.29	70.73
<b>Asian</b>	4,107,000	4.28	4,038,000	98.32	2,608,000	63.50	64.59
<b>AIAN</b>	1,179,000	1.23	1,175,000	99.66	872,000	73.96	74.21
<b>NHPI</b>	480,000	0.50	475,000	98.96	303,000	63.13	63.79
<b>Two or more races</b>	163,000	0.17	163,000	100	97,000	59.51	59.51
<b>Unknown</b>	18,690,000	19.47	18,240,000	97.59	11,890,000	63.62	65.19
<b>Total</b>	95,990,000	100	93,810,000	97.73	67,050,000	69.85	71.47

Source: T-MSIS Analytic Files (TAF) data (2016); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

Table A-2 functions as an especially useful tool for comparing the race/ethnicity distribution of beneficiaries missing and not missing race/ethnicity information in TAF across all states. For example, the first row for each state presents the race/ethnicity distribution among beneficiaries who are not missing on race/ethnicity for the original TAF Race variable. The second row for each state represents the race/ethnicity distribution based on Decennial/ACS data for beneficiaries missing on race/ethnicity in TAF. By comparing these two rows, it is possible to see how the race/ethnicity distribution differs between those who did and did not have race/ethnicity information in TAF for each state. For example, in California (CA), 57.19% of the beneficiaries were reported to be Hispanic. However, among those who were missing race/ethnicity, 34.17% were Hispanic based on Decennial/ACS data. Therefore, Hispanic beneficiaries were underrepresented among those with missing race/ethnicity information in TAF in California. The last two rows for each state provide the Supplemental and Prioritized Race approach race/ethnicity

distributions based on linked TAF and Decennial/ACS data. For example, when we supplement race/ethnicity information with Decennial/ACS data for those who were missing this information in TAF we found that 55.79% of Medicaid beneficiaries identified as Hispanic in California in 2016 (relative to 57.19%, originally). When race/ethnicity congruence is considered, as in the Prioritized Race approach, the percent of Hispanic beneficiaries in California increased to 58.75%.

**Table A-2. Comparing Racial and Ethnic Distribution for Beneficiaries with and without Race/Ethnicity Information in TAF Data across States**

State	Race variable	Hispanic	White	Black	Other
AK	Not missing race in TAF (TAF reported race)	3.07	41.14	4.14	51.66
	Missing TAF Race (Decennial/ACS reported race)	14.30	55.50	5.74	24.50
	Supplemental Race	3.51	41.71	4.20	50.58
	Prioritized Race	5.98	38.53	3.88	51.61
AL	Not missing race in TAF (TAF reported race)	2.67	53.63	42.09	1.61
	Missing TAF Race (Decennial/ACS reported race)	7.05	43.35	45.57	4.03
	Supplemental Race	4.36	49.65	43.44	2.54
	Prioritized Race	4.89	48.44	43.33	3.34
AR	Not missing race in TAF (TAF reported race)	28.64	69.87	0.00	1.49
	Missing TAF Race (Decennial/ACS reported race)	26.02	47.71	20.57	5.69
	Supplemental Race	28.15	65.69	3.88	2.28
	Prioritized Race	29.29	62.88	4.40	3.43
AZ	Not missing race in TAF (TAF reported race)	6.86	66.43	10.80	15.91
	Missing TAF Race (Decennial/ACS reported race)	68.66	21.92	4.12	5.31
	Supplemental Race	20.66	56.49	9.31	13.54
	Prioritized Race	34.44	43.23	8.46	13.87
CA	Not missing race in TAF (TAF reported race)	57.19	22.07	8.50	12.23
	Missing TAF Race (Decennial/ACS reported race)	34.17	35.70	9.48	20.65
	Supplemental Race	55.79	22.90	8.56	12.74
	Prioritized Race	58.75	19.77	8.07	13.41
CO	Not missing race in TAF (TAF reported race)	22.49	63.09	9.28	5.13
	Missing TAF Race (Decennial/ACS reported race)	43.13	43.56	6.08	7.23
	Supplemental Race	30.31	55.68	8.07	5.93
	Prioritized Race	32.84	52.47	7.85	6.84
CT	Not missing race in TAF (TAF reported race)	0	70.65	24.04	5.31
	Missing TAF Race (Decennial/ACS reported race)	57.23	26.20	12.11	4.46
	Supplemental Race	12.68	60.80	21.40	5.12
	Prioritized Race	22.12	50.96	20.37	6.54
DC	Not missing race in TAF (TAF reported race)	0	1.76	97.17	1.07
	Missing TAF Race (Decennial/ACS reported race)	34.30	20.49	40.15	5.05
	Supplemental Race	5.89	4.98	87.37	1.75
	Prioritized Race	6.91	4.79	85.52	2.78
DE	Not missing race in TAF (TAF reported race)	15.77	43.46	37.71	3.06
	Missing TAF Race (Decennial/ACS reported race)	5.00	50.00	30.00	20.00
	Supplemental Race	15.77	43.46	37.71	3.06
	Prioritized Race	17.82	39.62	36.92	5.65

<b>State</b>	<b>Race variable</b>	<b>Hispanic</b>	<b>White</b>	<b>Black</b>	<b>Other</b>
<b>FL</b>	Not missing race in TAF (TAF reported race)	36.23	34.40	27.77	1.59
	Missing TAF Race (Decennial/ACS reported race)	32.44	33.65	25.67	8.23
	Supplemental Race	35.90	34.33	27.59	2.18
	Prioritized Race	37.49	32.31	26.93	3.27
<b>GA</b>	Not missing race in TAF (TAF reported race)	2.57	44.87	49.59	2.96
	Missing TAF Race (Decennial/ACS reported race)	9.50	33.68	46.83	9.99
	Supplemental Race	3.04	44.12	49.41	3.43
	Prioritized Race	8.59	38.00	48.87	4.54
<b>HI</b>	Not missing race in TAF (TAF reported race)	2.64	23.79	1.96	71.62
	Missing TAF Race (Decennial/ACS reported race)	12.83	30.43	1.96	54.78
	Supplemental Race	3.88	24.59	1.96	69.57
	Prioritized Race	11.63	18.13	1.42	68.82
<b>IA</b>	Not missing race in TAF (TAF reported race)	70.32	2.35	10.89	16.44
	Missing TAF Race (Decennial/ACS reported race)	6.76	75.70	9.21	8.34
	Supplemental Race	58.18	16.36	10.57	14.89
	Prioritized Race	63.37	16.30	10.08	10.25
<b>ID</b>	Not missing race in TAF (TAF reported race)	0.03	98.06	0.00	1.91
	Missing TAF Race (Decennial/ACS reported race)	10.00	85.00	0.85	3.80
	Supplemental Race	0.05	98.03	0.00	1.91
	Prioritized Race	13.18	81.35	0.70	4.78
<b>IL</b>	Not missing race in TAF (TAF reported race)	23.07	42.38	30.62	3.93
	Missing TAF Race (Decennial/ACS reported race)	24.26	41.49	21.98	12.27
	Supplemental Race	23.12	42.34	30.27	4.26
	Prioritized Race	26.05	38.27	29.95	5.73
<b>IN</b>	Not missing race in TAF (TAF reported race)	9.85	66.83	21.03	2.29
	Missing TAF Race (Decennial/ACS reported race)	19.31	56.74	15.37	8.58
	Supplemental Race	10.15	66.51	20.85	2.48
	Prioritized Race	11.83	63.30	20.36	4.51
<b>KS</b>	Not missing race in TAF (TAF reported race)	100	0	0	0
	Missing TAF Race (Decennial/ACS reported race)	6.16	71.09	13.62	9.13
	Supplemental Race	31.23	52.10	9.98	6.69
	Prioritized Race	31.23	52.10	9.98	6.69
<b>KY</b>	Not missing race in TAF (TAF reported race)	3.62	82.29	12.92	1.17
	Missing TAF Race (Decennial/ACS reported race)	4.46	77.33	12.94	5.27
	Supplemental Race	3.74	81.60	12.92	1.74
	Prioritized Race	4.65	79.14	13.00	3.22
<b>LA</b>	Not missing race in TAF (TAF reported race)	6.61	41.38	50.87	1.14
	Missing TAF Race (Decennial/ACS reported race)	5.47	42.08	46.27	6.18
	Supplemental Race	6.38	41.53	49.92	2.18
	Prioritized Race	7.47	39.32	49.85	3.36
<b>MA</b>	Not missing race in TAF (TAF reported race)	9.86	64.48	17.54	8.12
	Missing TAF Race (Decennial/ACS reported race)	34.25	46.05	10.53	9.17
	Supplemental Race	18.50	57.95	15.06	8.49
	Prioritized Race	20.52	55.54	14.57	9.37
<b>MD</b>	Not missing race in TAF (TAF reported race)	16.15	31.56	45.97	6.32
	Missing TAF Race (Decennial/ACS reported race)	9.24	36.64	43.38	10.73
	Supplemental Race	15.34	32.15	45.67	6.84
	Prioritized Race	16.12	30.60	44.91	8.38

State	Race variable	Hispanic	White	Black	Other
ME	Not missing race in TAF (TAF reported race)	1.97	91.92	4.14	1.98
	Missing TAF Race (Decennial/ACS reported race)	2.43	90.61	2.62	4.34
	Supplemental Race	2.00	91.84	4.04	2.12
	Prioritized Race	2.89	88.53	4.18	4.39
MI	Not missing race in TAF (TAF reported race)	6.71	62.44	28.98	1.87
	Missing TAF Race (Decennial/ACS reported race)	6.03	52.67	24.52	16.78
	Supplemental Race	6.67	61.84	28.71	2.79
	Prioritized Race	8.60	57.81	28.18	5.41
MN	Not missing race in TAF (TAF reported race)	7.72	59.93	20.49	11.86
	Missing TAF Race (Decennial/ACS reported race)	12.74	57.71	16.72	12.83
	Supplemental Race	8.36	59.64	20.01	11.98
	Prioritized Race	9.73	57.65	18.75	13.87
MO	Not missing race in TAF (TAF reported race)	10.52	65.40	22.56	1.52
	Missing TAF Race (Decennial/ACS reported race)	5.23	65.41	24.05	5.32
	Supplemental Race	8.35	65.40	23.17	3.08
	Prioritized Race	9.17	63.10	23.04	4.69
MS	Not missing race in TAF (TAF reported race)	2.04	42.91	53.62	1.43
	Missing TAF Race (Decennial/ACS reported race)	2.82	37.32	57.37	2.49
	Supplemental Race	2.64	38.55	56.54	2.26
	Prioritized Race	2.87	38.15	56.45	2.53
MT	Not missing race in TAF (TAF reported race)	3.99	75.14	0.86	20.01
	Missing TAF Race (Decennial/ACS reported race)	5.51	86.37	1.27	6.85
	Supplemental Race	4.10	75.98	0.89	19.02
	Prioritized Race	6.06	72.20	0.85	20.89
NC	Not missing race in TAF (TAF reported race)	13.19	44.21	36.31	6.29
	Missing TAF Race (Decennial/ACS reported race)	20.68	44.47	25.66	9.20
	Supplemental Race	13.24	44.21	36.25	6.30
	Prioritized Race	15.86	40.63	35.52	7.99
ND	Not missing race in TAF (TAF reported race)	4.68	62.05	8.63	24.65
	Missing TAF Race (Decennial/ACS reported race)	3.06	84.50	2.04	10.40
	Supplemental Race	4.65	62.39	8.53	24.43
	Prioritized Race	6.66	59.39	8.44	25.51
NE	Not missing race in TAF (TAF reported race)	0	0	0	0
	Missing TAF Race (Decennial/ACS reported race)	19.14	60.72	11.96	8.18
	Supplemental Race	19.14	60.72	11.96	8.18
	Prioritized Race	19.14	60.72	11.96	8.18
NH	Not missing race in TAF (TAF reported race)	5.65	88.47	2.40	3.48
	Missing TAF Race (Decennial/ACS reported race)	7.72	79.52	4.08	8.68
	Supplemental Race	5.74	88.09	2.47	3.70
	Prioritized Race	7.52	84.32	2.63	5.52
NJ	Not missing race in TAF (TAF reported race)	27.46	40.53	26.74	5.26
	Missing TAF Race (Decennial/ACS reported race)	24.11	42.61	13.43	19.86
	Supplemental Race	27.15	40.73	25.51	6.61
	Prioritized Race	33.43	34.16	24.53	7.87
NM	Not missing race in TAF (TAF reported race)	54.35	28.02	2.16	15.47
	Missing TAF Race (Decennial/ACS reported race)	30.40	56.22	3.75	9.63
	Supplemental Race	54.04	28.39	2.18	15.39
	Prioritized Race	60.19	21.92	2.06	15.83

State	Race variable	Hispanic	White	Black	Other
NV	Not missing race in TAF (TAF reported race)	34.51	38.05	20.54	6.90
	Missing TAF Race (Decennial/ACS reported race)	36.33	21.70	10.65	31.33
	Supplemental Race	34.55	37.61	20.28	7.56
	Prioritized Race	38.17	33.03	19.28	9.52
NY	Not missing race in TAF (TAF reported race)	16.78	43.35	24.29	15.57
	Missing TAF Race (Decennial/ACS reported race)	49.00	27.08	14.51	9.41
	Supplemental Race	23.13	40.15	22.37	14.36
	Prioritized Race	26.09	36.94	21.45	15.52
OH	Not missing race in TAF (TAF reported race)	4.15	64.59	29.13	2.14
	Missing TAF Race (Decennial/ACS reported race)	5.31	76.46	13.22	5.02
	Supplemental Race	4.22	65.36	28.09	2.33
	Prioritized Race	5.82	62.03	27.34	4.81
OK	Not missing race in TAF (TAF reported race)	17.41	51.32	13.48	17.79
	Missing TAF Race (Decennial/ACS reported race)	8.43	54.79	16.07	20.71
	Supplemental Race	17.11	51.43	13.56	17.89
	Prioritized Race	18.76	47.11	12.40	21.74
OR	Not missing race in TAF (TAF reported race)	31.18	60.06	3.05	5.70
	Missing TAF Race (Decennial/ACS reported race)	50.43	38.36	2.13	9.09
	Supplemental Race	34.01	56.88	2.92	6.20
	Prioritized Race	36.79	52.66	2.77	7.78
PA	Not missing race in TAF (TAF reported race)	12.54	59.03	25.05	3.38
	Missing TAF Race (Decennial/ACS reported race)	26.23	48.47	14.03	11.26
	Supplemental Race	13.01	58.66	24.67	3.65
	Prioritized Race	14.93	55.34	24.22	5.51
RI	Not missing race in TAF (TAF reported race)	0	0	0	0
	Missing TAF Race (Decennial/ACS reported race)	25.11	59.17	9.14	6.58
	Supplemental Race	25.11	59.17	9.14	6.58
	Prioritized Race	25.11	59.17	9.14	6.58
SC	Not missing race in TAF (TAF reported race)	5.64	45.39	47.56	1.42
	Missing TAF Race (Decennial/ACS reported race)	8.10	46.64	39.74	5.52
	Supplemental Race	6.12	45.63	46.02	2.22
	Prioritized Race	7.29	43.37	45.78	3.55
SD	Not missing race in TAF (TAF reported race)	4.86	61.53	3.67	29.94
	Missing TAF Race (Decennial/ACS reported race)	5.00	20.00	10.00	70.00
	Supplemental Race	4.86	61.51	3.67	29.96
	Prioritized Race	7.10	54.27	3.99	34.64
TN	Not missing race in TAF (TAF reported race)	0	0	0	0
	Missing TAF Race (Decennial/ACS reported race)	5.71	63.37	27.37	3.56
	Supplemental Race	5.71	63.37	27.37	3.56
	Prioritized Race	5.71	63.37	27.37	3.56
TX	Not missing race in TAF (TAF reported race)	58.05	22.53	16.28	3.14
	Missing TAF Race (Decennial/ACS reported race)	26.13	34.39	31.07	8.40
	Supplemental Race	56.17	23.22	17.15	3.45
	Prioritized Race	58.65	20.53	16.71	4.11
UT	Not missing race in TAF (TAF reported race)	25.94	68.45	2.91	2.70
	Missing TAF Race (Decennial/ACS reported race)	14.13	62.68	2.18	21.01
	Supplemental Race	24.08	67.54	2.79	5.58
	Prioritized Race	28.49	61.53	2.61	7.38

State	Race variable	Hispanic	White	Black	Other
VA	Not missing race in TAF (TAF reported race)	3.13	53.94	37.97	4.96
	Missing TAF Race (Decennial/ACS reported race)	39.63	31.77	18.73	9.87
	Supplemental Race	3.52	53.70	37.76	5.02
	Prioritized Race	8.72	46.53	37.39	7.36
VT	Not missing race in TAF (TAF reported race)	0.55	94.71	2.59	2.15
	Missing TAF Race (Decennial/ACS reported race)	2.07	91.99	1.50	4.44
	Supplemental Race	0.76	94.32	2.44	2.48
	Prioritized Race	1.98	90.69	2.60	4.74
WA	Not missing race in TAF (TAF reported race)	21.79	58.42	7.52	12.26
	Missing TAF Race (Decennial/ACS reported race)	20.32	54.02	5.78	19.88
	Supplemental Race	21.70	58.16	7.41	12.72
	Prioritized Race	23.61	54.43	6.84	15.12
WI	Not missing race in TAF (TAF reported race)	13.02	62.40	18.97	5.61
	Missing TAF Race (Decennial/ACS reported race)	7.36	56.96	19.86	15.82
	Supplemental Race	12.45	61.85	19.06	6.64
	Prioritized Race	13.65	59.93	18.58	7.84
WV	Not missing race in TAF (TAF reported race)	0.00	93.54	4.61	1.85
	Missing TAF Race (Decennial/ACS reported race)	2.18	89.44	5.45	2.92
	Supplemental Race	0.35	92.89	4.74	2.02
	Prioritized Race	1.32	90.11	5.15	3.42
WY	Not missing race in TAF (TAF reported race)	1.31	85.44	2.46	10.79
	Missing TAF Race (Decennial/ACS reported race)	42.05	51.81	1.08	5.06
	Supplemental Race	7.75	80.13	2.24	9.88
	Prioritized Race	12.51	74.23	1.90	11.36

Source: T-MSIS Analytic Files (TAF) data (2016); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: All rows total 100. Blank cells are indicative of a state missing information for a particular race/ethnicity group. DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

Table A-3 is similar to Table 5 but provides the race/ethnicity distribution for the detailed race/ethnicity variable. Among the newly added race/ethnicity groups (Asian, AIAN, NHPI, two or more races), a lower percentage of Asian, AIAN, and NHPI beneficiaries were missing TAF race/ethnicity relative to those not missing race in TAF. However, beneficiaries who identified as two or more races were nearly much more likely to be missing on TAF race/ethnicity relative to not missing race/ethnicity in TAF (0.21% and 3.31%, respectively).

**Table A-3. Comparing Detailed Race/Ethnicity Distribution for Beneficiaries with and without Race/Ethnicity Information in TAF data**

	Not missing race in TAF (TAF reported race)		Missing TAF race (Decennial/ACS reported race)	
	N	Column %	N	Column %
	(1)	(2)	(3)	(4)
<b>Hispanic</b>	21,120,000	27.32	2,901,000	24.46
<b>White</b>	34,060,000	44.06	5,466,000	46.09
<b>Black</b>	16,190,000	20.94	2,435,000	20.53
<b>Asian</b>	4,107,000	5.31	554,000	4.67
<b>AIAN</b>	1,179,000	1.53	89,500	0.75
<b>NHPI</b>	480,000	0.62	28,000	0.24
<b>Two or more races</b>	163,000	0.21	392,000	3.31
<b>Total</b>	77,300,000	100	11,860,000	100

Source: T-MSIS Analytic Files (TAF) data (2016); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

Table A-4 provides the distribution of Medicaid beneficiaries across states for all race/ethnicity groups. The PIK rate was relatively high (>98%) for most states (n=46) (Column 4). The PIK rate was lowest in California (92.22%), Massachusetts (95.44%), Arizona (96.42%), Oregon (97.34%), and New York (97.92%). The percent of beneficiaries linked to Decennial/ACS data varied across states. Texas (61.27%), Utah (65.16%), and Georgia (65.18%) had the lowest linkage rates while Vermont (82.38%), Maine (80.06%), and West Virginia (78.53%) had the highest linkage rates (Column 6). Conditional upon receiving a PIK, linkage rates for each state increased slightly (Column 7). There was not a strong association between PIK and linkage rates ( $r = 0.31$ ). For example, Florida and Minnesota shared a PIK rate of 99.42% but had different linkage rates (66.17% and 75.34%, respectively).

**Table A-4. State Distribution Based on 2016 TAF Data**

	All observations		Observations with PIKs		Observations Linked to Decennial/ACS		
	N	Column %	N	Row %	N	Linkage Rate (%)	Linkage Rate (%), Conditional on PIK
State	(1)	(2)	(3)	(4)	(5)	(6)	(7)
AK	209,000	0.22	209,000	100	156,000	74.64	74.64
AL	1,335,000	1.39	1,321,000	98.95	926,000	69.36	70.10
AR	1,165,000	1.21	1,161,000	99.66	888,000	76.22	76.49
AZ	2,377,000	2.48	2,292,000	96.42	1,638,000	68.91	71.47
CA	17,620,000	18.36	16,250,000	92.22	11,540,000	65.49	71.02
CO	1,649,000	1.72	1,640,000	99.45	1,232,000	74.71	75.12
CT	1,063,000	1.11	1,058,000	99.53	809,000	76.11	76.47
DC	283,000	0.29	281,000	99.29	204,000	72.08	72.60
DE	295,000	0.31	291,000	98.64	211,000	71.53	72.51
FL	5,188,000	5.40	5,158,000	99.42	3,433,000	66.17	66.56
GA	2,496,000	2.60	2,474,000	99.12	1,627,000	65.18	65.76
HI	423,000	0.44	422,000	99.76	290,000	68.56	68.72
IA	806,000	0.84	793,000	98.39	609,000	75.56	76.80
ID	365,000	0.38	364,000	99.73	248,000	67.95	68.13
IL	3,744,000	3.90	3,722,000	99.41	2,757,000	73.64	74.07
IN	1,841,000	1.92	1,820,000	98.86	1,362,000	73.98	74.84
KS	519,000	0.54	517,000	99.61	357,000	68.79	69.05
KY	1,655,000	1.72	1,645,000	99.40	1,272,000	76.86	77.33
LA	1,747,000	1.82	1,742,000	99.71	1,282,000	73.38	73.59
MA	2,238,000	2.33	2,136,000	95.44	1,522,000	68.01	71.25
MD	1,536,000	1.60	1,523,000	99.15	1,085,000	70.64	71.24
ME	331,000	0.34	330,000	99.70	265,000	80.06	80.30
MI	2,923,000	3.05	2,902,000	99.28	2,252,000	77.04	77.60
MN	1,391,000	1.45	1,383,000	99.42	1,048,000	75.34	75.78
MO	1,232,000	1.28	1,227,000	99.59	867,000	70.37	70.66
MS	911,000	0.95	910,000	99.89	646,000	70.91	70.99
MT	277,000	0.29	276,000	99.64	208,000	75.09	75.36
NC	2,420,000	2.52	2,407,000	99.46	1,683,000	69.55	69.92
ND	124,000	0.13	124,000	100	89,500	72.18	72.18
NE	304,000	0.32	304,000	100	201,000	66.12	66.12
NH	261,000	0.27	260,000	99.62	202,000	77.39	77.69
NJ	2,119,000	2.21	2,106,000	99.39	1,473,000	69.51	69.94
NM	972,000	1.01	968,000	99.59	733,000	75.41	75.72
NV	841,000	0.88	830,000	98.69	584,000	69.44	70.36
NY	7,876,000	8.21	7,712,000	97.92	5,218,000	66.25	67.66
OH	3,495,000	3.64	3,480,000	99.57	2,687,000	76.88	77.21
OK	1,025,000	1.07	1,021,000	99.61	719,000	70.15	70.42
OR	1,427,000	1.49	1,389,000	97.34	1,069,000	74.91	76.96
PA	3,325,000	3.46	3,311,000	99.58	2,516,000	75.67	75.99
RI	358,000	0.37	355,000	99.16	263,000	73.46	74.08

	All observations		Observations with PIKs		Observations Linked to Decennial/ACS		
	N	Column %	N	Row %	N	Linkage Rate (%)	Linkage Rate (%), Conditional on PIK
<b>State</b>	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<b>SC</b>	1,404,000	1.46	1,395,000	99.36	1,015,000	72.29	72.76
<b>SD</b>	155,000	0.16	155,000	100	104,000	67.1	67.10
<b>TN</b>	1,796,000	1.87	1,789,000	99.61	1,306,000	72.72	73.00
<b>TX</b>	5,952,000	6.20	5,878,000	98.76	3,647,000	61.27	62.04
<b>UT</b>	442,000	0.46	439,000	99.32	288,000	65.16	65.60
<b>VA</b>	1,428,000	1.49	1,420,000	99.44	987,000	69.12	69.51
<b>VT</b>	227,000	0.24	226,000	99.56	187,000	82.38	82.74
<b>WA</b>	2,201,000	2.29	2,188,000	99.41	1,625,000	73.83	74.27
<b>WI</b>	1,419,000	1.48	1,415,000	99.72	1,105,000	77.87	78.09
<b>WV</b>	694,000	0.72	693,000	99.86	545,000	78.53	78.64
<b>WY</b>	103,000	0.11	103,000	100	73,000	70.87	70.87
<b>Total</b>	95,990,000	100	93,810,000	97.73	67,050,000	69.85	71.47

Source: T-MSIS Analytic Files (TAF) data (2016); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: Column 7 percentage was calculated as the number of observations that were linked to Decennial/ACS data divided by the number of observations that received a PIK (Column 3). DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

Table A-5 provides the state-level percentages in table format from the maps included in the data linkage portion of our analysis (Figure 1 - Figure 3).

**Table A-5. State-Level Percentages for Figure 1 – Figure 3**

	<b>Figure 1</b>	<b>Figure 2</b>	<b>Figure 3</b>
	Percent missing race/ethnicity	Percent missing populated by linkage	Percent missing after linkage
<b>State</b>	(1)	(2)	(3)
<b>AK</b>	5.78	67.71	1.87
<b>AL</b>	50.34	62.19	19.03
<b>AR</b>	25.40	68.26	8.06
<b>AZ</b>	31.09	63.74	11.27
<b>CA</b>	10.05	58.11	4.21
<b>CO</b>	46.34	70.67	13.59
<b>CT</b>	30.04	66.28	10.13
<b>DC</b>	24.95	62.42	9.37
<b>DE</b>	0.01	80.00	0.00
<b>FL</b>	13.85	60.16	5.52
<b>GA</b>	9.80	66.21	3.31
<b>HI</b>	20.53	53.47	9.55
<b>IA</b>	26.53	65.38	9.19
<b>ID</b>	0.46	42.60	0.26
<b>IL</b>	5.50	71.34	1.58
<b>IN</b>	4.28	71.75	1.21
<b>KS</b>	79.45	70.98	23.05
<b>KY</b>	17.54	76.22	4.17
<b>LA</b>	26.30	72.81	7.15
<b>MA</b>	47.60	60.39	18.85
<b>MD</b>	17.09	64.14	6.13
<b>ME</b>	8.94	66.29	3.01
<b>MI</b>	9.31	63.97	3.35
<b>MN</b>	18.85	63.32	6.91
<b>MO</b>	51.86	64.50	18.41
<b>MS</b>	83.52	69.56	25.42
<b>MT</b>	11.48	62.96	4.25
<b>NC</b>	1.04	59.88	0.42
<b>ND</b>	1.71	88.00	0.21
<b>NE</b>	100	65.98	34.02
<b>NH</b>	7.71	53.00	3.62
<b>NJ</b>	11.89	75.37	2.93
<b>NM</b>	1.84	70.89	0.54
<b>NV</b>	4.30	61.55	1.65
<b>NY</b>	30.46	55.97	13.41
<b>OH</b>	9.46	66.57	3.16
<b>OK</b>	4.86	65.73	1.67
<b>OR</b>	20.95	64.92	7.35
<b>PA</b>	5.17	65.18	1.80
<b>RI</b>	100	73.26	26.74

	<b>Figure 1</b>	<b>Figure 2</b>	<b>Figure 3</b>
	Percent missing race/ethnicity	Percent missing populated by linkage	Percent missing after linkage
<b>State</b>	(1)	(2)	(3)
<b>SC</b>	28.90	60.03	11.55
<b>SD</b>	0.08	61.00	0.03
<b>TN</b>	100	72.58	27.42
<b>TX</b>	12.12	45.35	6.62
<b>UT</b>	33.66	36.82	21.26
<b>VA</b>	1.88	57.24	0.80
<b>VT</b>	17.57	77.90	3.88
<b>WA</b>	10.34	55.56	4.59
<b>WI</b>	15.12	63.38	5.54
<b>WV</b>	20.28	74.45	5.18
<b>WY</b>	24.50	57.82	10.33

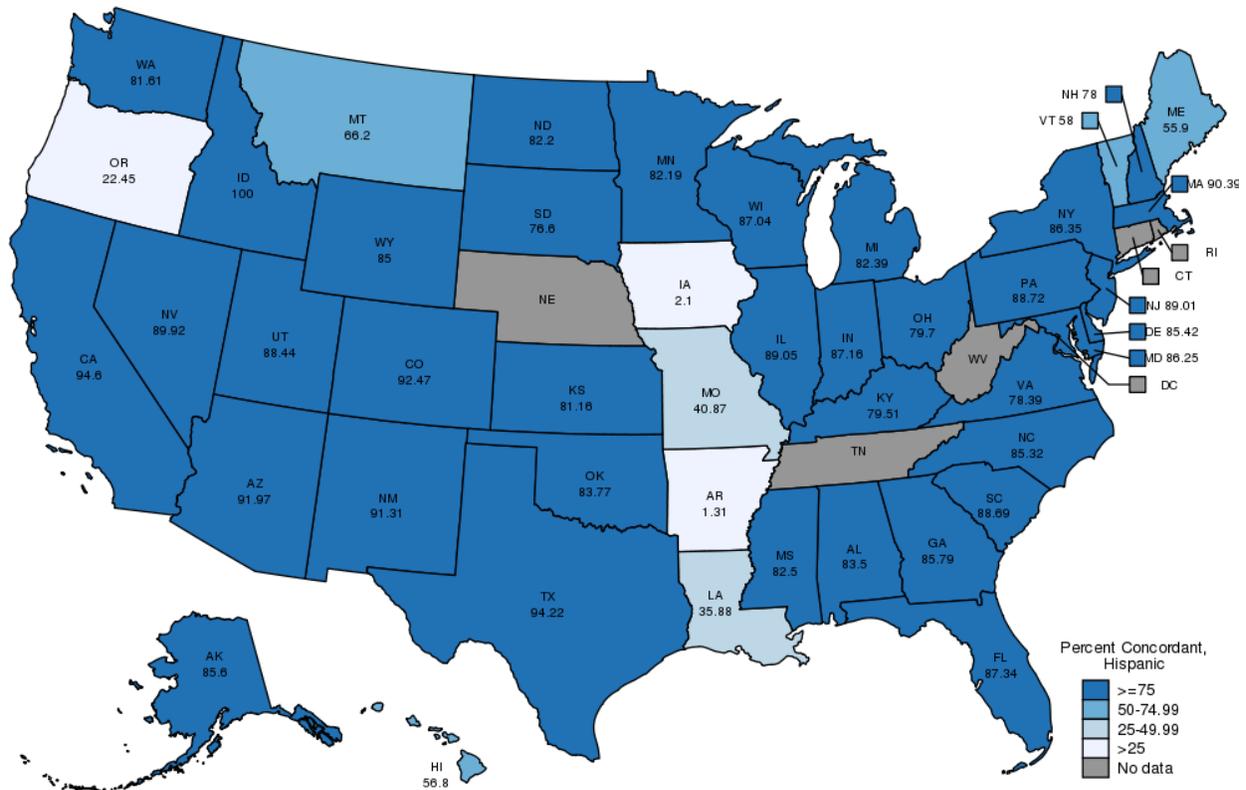
Source: T-MSIS Analytic Files (TAF) data (2016); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

Appendix B. Race/Ethnicity Concordance Supplement

Figure B-1 through Figure B-4 show the percent of observations with race/ethnicity concordance for each of the four race/ethnicity categories (i.e., Hispanic; White; Black; and other race) between the two data sources for all states. Among those with race/ethnicity in both sources, the percent of Medicaid beneficiaries with race/ethnicity concordance was defined as the number of beneficiaries who had concordant race/ethnicity in TAF and Decennial/ACS data divided by the number of beneficiaries reporting that race/ethnicity in TAF. We also provide a table with all these values listed by state for easy reference (Table B-1). The average concordance rate was lowest for other race (81.38%) and Hispanic (86.16%) beneficiaries and highest for Black (89.87%) and White (88.08%) beneficiaries.

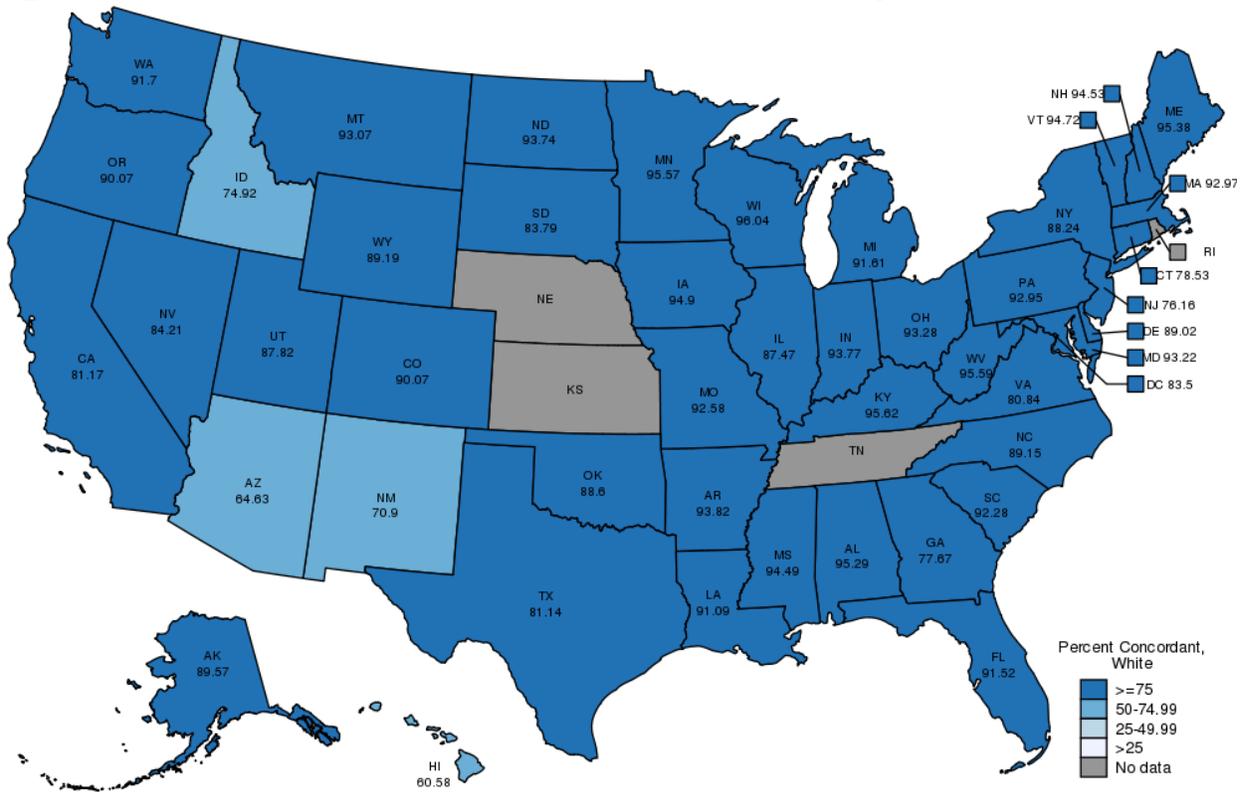
**Figure B-1. Percent of Medicaid Beneficiaries with Race/Ethnicity Concordance, Hispanic**



Source: T-MSIS Analytic Files (TAF) data (2016); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: Race/ethnicity information was not available in TAF data for Nebraska, Rhode Island, and Tennessee. Connecticut and DC were missing entirely on Hispanic ethnicity information for beneficiaries. Information for West Virginia has been suppressed due to a small denominator. DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy

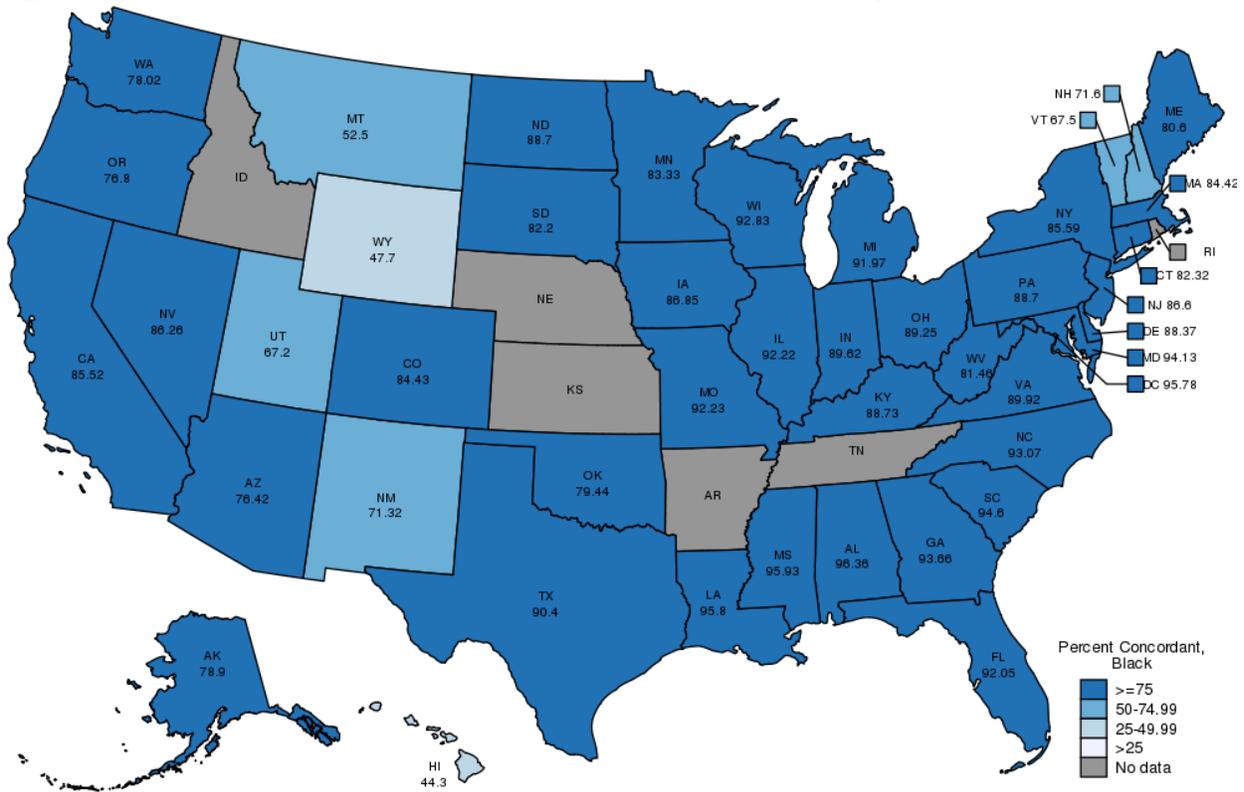
**Figure B-2. Percent of Medicaid Beneficiaries with Race/Ethnicity Concordance, White**



Source: T-MSIS Analytic Files (TAF) data (2016); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: Race/ethnicity information was not available in TAF data for Nebraska, Rhode Island, and Tennessee. Kansas was missing entirely for White beneficiaries. DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy

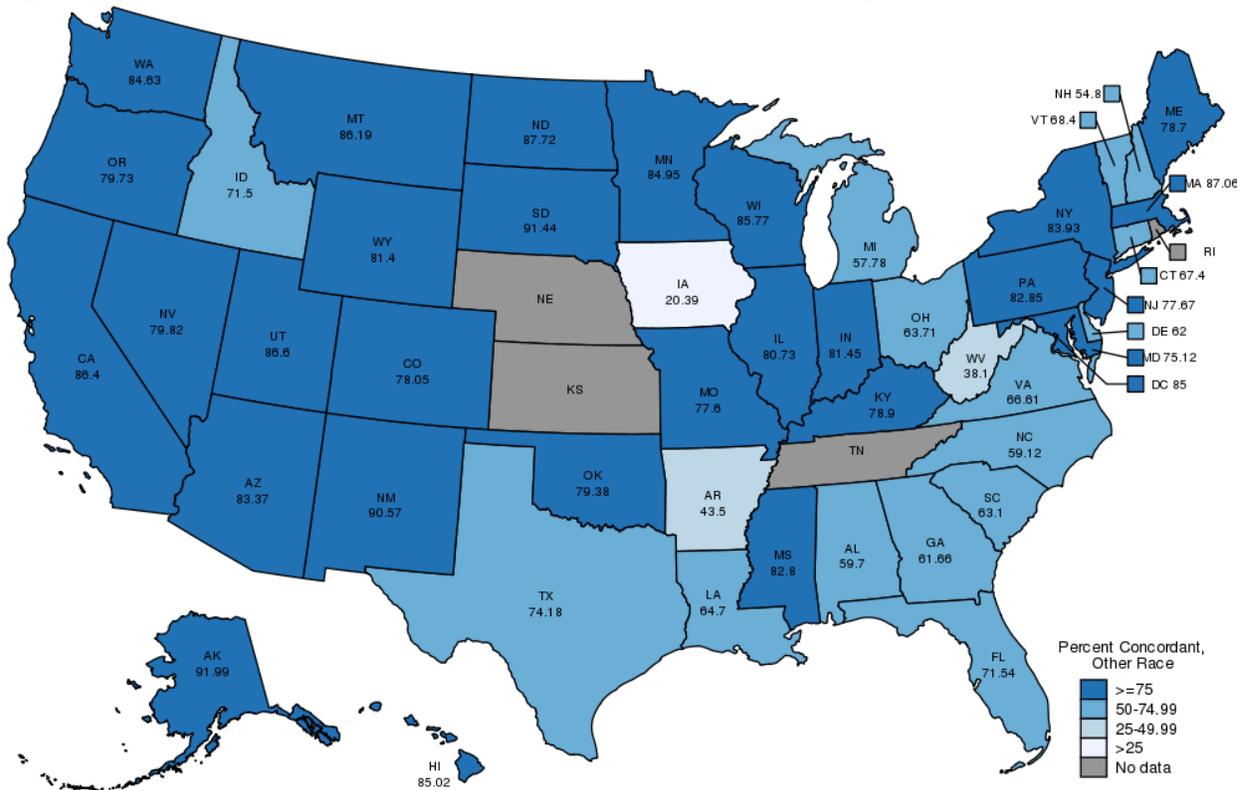
**Figure B-3. Percent of Medicaid Beneficiaries with Race/Ethnicity Concordance, Black**



Source: T-MSIS Analytic Files (TAF) data (2016); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: Race/ethnicity information was not available in TAF data for Nebraska, Rhode Island, and Tennessee. Kansas was missing entirely for White beneficiaries. Information for Arkansas and Idaho has been suppressed due to small denominators. DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

**Figure B-4. Percent of Medicaid Beneficiaries with Race/Ethnicity Concordance, Other Race**



Source: T-MSIS Analytic Files (TAF) data (2016); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: Race/ethnicity information was not available in TAF data for Nebraska, Rhode Island, and Tennessee. Kansas was missing entirely for White beneficiaries. DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

Table B-1 provides the state-level percentages in table format from the race/ethnicity concordance maps (Figure B-1 through Figure B-4).

**Table B-1. Percent of Medicaid Beneficiaries with Race/Ethnicity Concordance**

	<b>Figure B-1</b>	<b>Figure B-2</b>	<b>Figure B-3</b>	<b>Figure B-4</b>
	Percent concordant, Hispanic	Percent concordant, White	Percent concordant, Black	Percent concordant, Other
<b>State</b>	(1)	(2)	(3)	(4)
<b>AK</b>	85.60	89.57	78.90	91.99
<b>AL</b>	83.50	95.29	96.36	59.70
<b>AR</b>	1.31	93.82		43.50
<b>AZ</b>	91.97	64.63	76.42	83.37
<b>CA</b>	94.60	81.17	85.52	86.40
<b>CO</b>	92.47	90.07	84.43	78.05
<b>CT</b>		78.53	82.32	67.40
<b>DC</b>		83.50	95.78	85.00
<b>DE</b>	85.42	89.02	88.37	62.00
<b>FL</b>	87.34	91.52	92.05	71.54
<b>GA</b>	85.79	77.67	93.66	61.66
<b>HI</b>	56.80	60.58	44.30	85.02
<b>IA</b>	2.10	94.90	86.85	20.39
<b>ID</b>	100	74.92		71.50
<b>IL</b>	89.05	87.47	92.22	80.73
<b>IN</b>	87.16	93.77	89.62	81.45
<b>KS</b>	81.16			
<b>KY</b>	79.51	95.62	88.73	78.90
<b>LA</b>	35.88	91.09	95.80	64.70
<b>MA</b>	90.39	92.97	84.42	87.06
<b>MD</b>	86.25	93.22	94.13	75.12
<b>ME</b>	55.90	95.38	80.60	78.70
<b>MI</b>	82.39	91.61	91.97	57.78
<b>MN</b>	82.19	95.57	83.33	84.95
<b>MO</b>	40.87	92.58	92.23	77.60
<b>MS</b>	82.50	94.49	95.93	82.80
<b>MT</b>	66.20	93.07	52.50	86.19
<b>NC</b>	85.32	89.15	93.07	59.12
<b>ND</b>	82.20	93.74	88.70	87.72
<b>NE</b>				
<b>NH</b>	78.00	94.53	71.60	54.80
<b>NJ</b>	89.01	76.16	86.60	77.67
<b>NM</b>	91.31	70.90	71.32	90.57
<b>NV</b>	89.92	84.21	86.26	79.82
<b>NY</b>	86.35	88.24	85.59	83.93
<b>OH</b>	79.70	93.28	89.25	63.71
<b>OK</b>	83.77	88.60	79.44	79.38
<b>OR</b>	22.45	90.07	76.80	79.73
<b>PA</b>	88.72	92.95	88.70	82.85
<b>RI</b>				

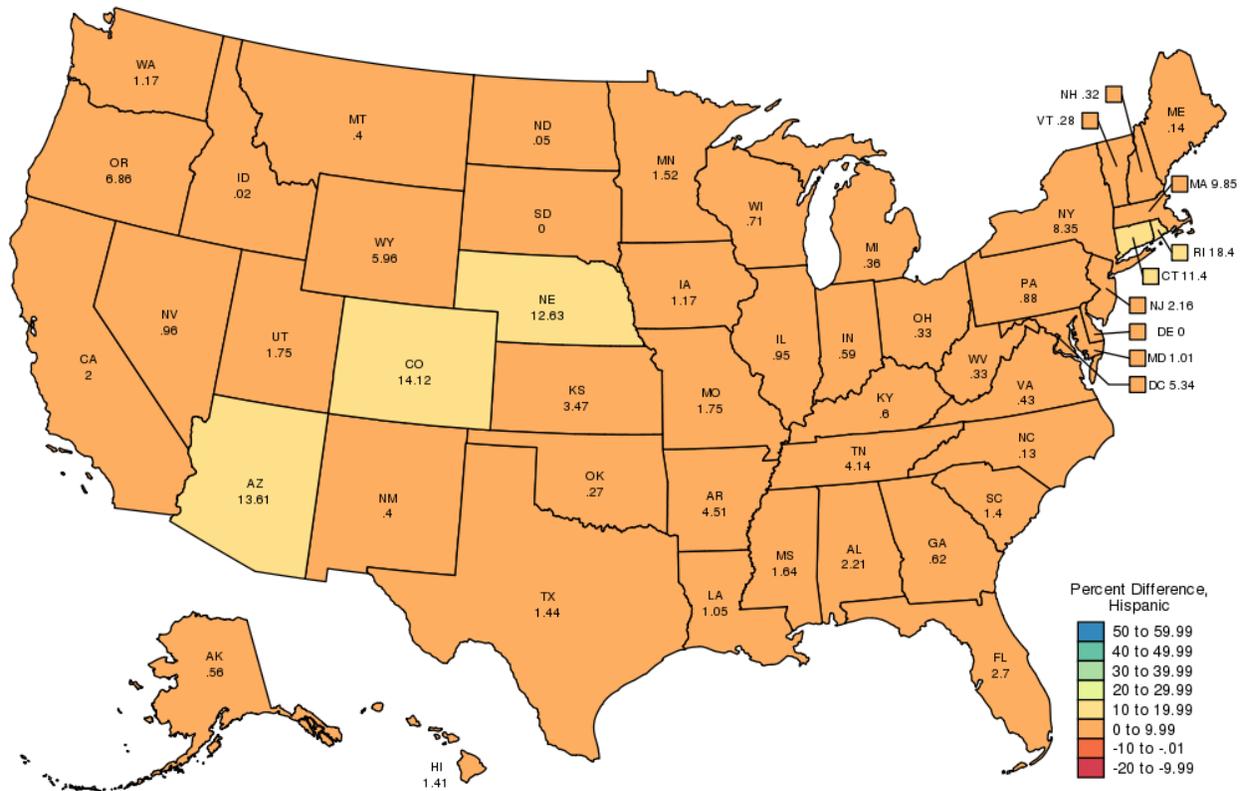
	<b>Figure B-1</b>	<b>Figure B-2</b>	<b>Figure B-3</b>	<b>Figure B-4</b>
	Percent concordant, Hispanic	Percent concordant, White	Percent concordant, Black	Percent concordant, Other
<b>State</b>	(1)	(2)	(3)	(4)
<b>SC</b>	88.69	92.28	94.60	63.10
<b>SD</b>	76.60	83.79	82.20	91.44
<b>TN</b>				
<b>TX</b>	94.22	81.14	90.40	74.18
<b>UT</b>	88.44	87.82	67.20	86.60
<b>VA</b>	78.39	80.84	89.92	66.61
<b>VT</b>	58.00	94.72	67.50	68.40
<b>WA</b>	81.61	91.70	78.02	84.63
<b>WI</b>	87.04	96.04	92.83	85.77
<b>WV</b>		95.59	81.46	38.10
<b>WY</b>	85.00	89.19	47.70	81.40

Source: T-MSIS Analytic Files (TAF) data (2016); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

Figure B-5 through Figure B-8 provide information on the percentage point difference between TAF Race and Supplemental Race for each of the four race/ethnic groups across all states. Nebraska, Rhode Island, and Tennessee were missing entirely on race/ethnicity information in TAF data, so these states started at 0%. All maps used the same scale and color theme to make comparisons across racial/ethnic groups easier.

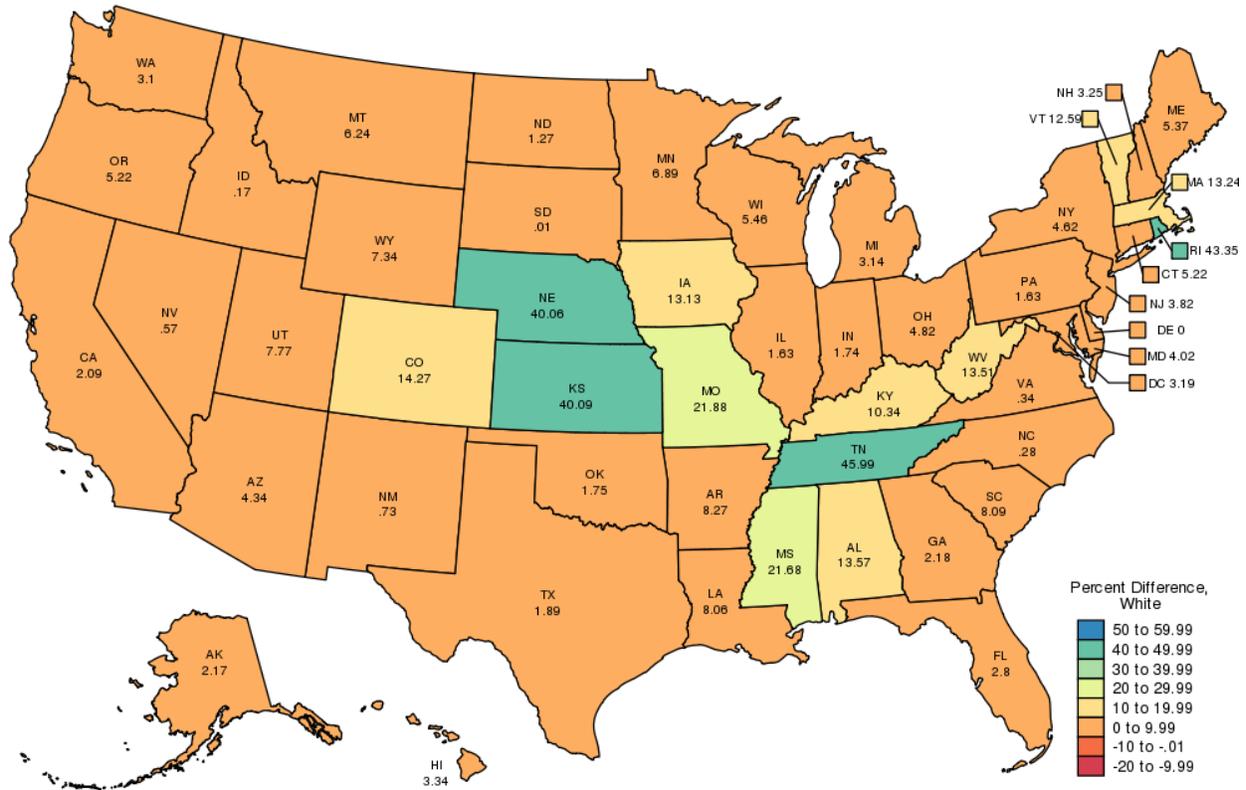
**Figure B-5. Percentage Point Difference in Race/Ethnicity Distribution Between TAF Race and Supplemental Race, Hispanic**



Source: T-MSIS Analytic Files (TAF) data (2016); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: Percentage point difference is measured as the percent of observations identified as a particular race/ethnicity in Supplemental Race minus the percent in TAF Race. Therefore, a positive value denotes an increase in the race/ethnic category when using the Supplemental Race approach. Connecticut, DC, Nebraska, Rhode Island, and Tennessee reported no beneficiaries with a Hispanic identity in TAF and started at 0% for this comparison map. DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy

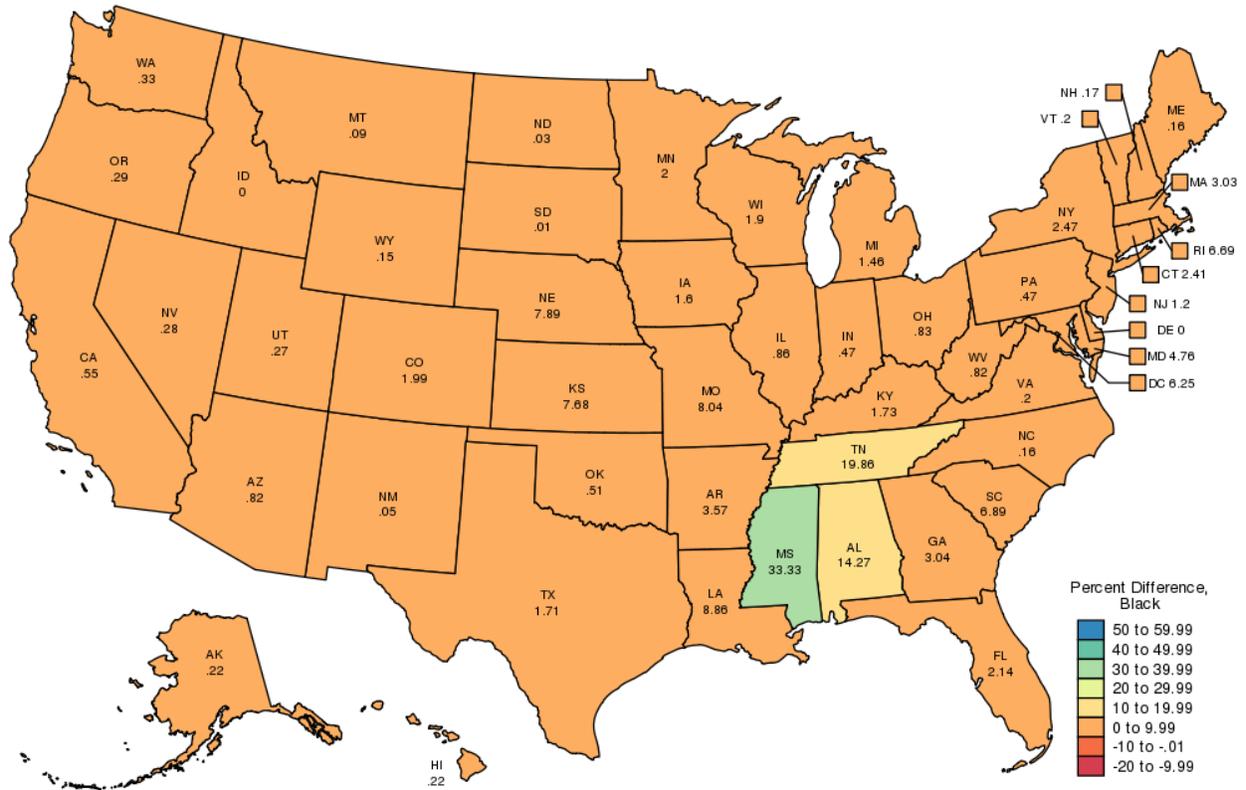
**Figure B-6. Percentage Point Difference in Race/Ethnicity Distribution Between TAF Race and Supplemental Race, White**



Source: T-MSIS Analytic Files (TAF) data (2016); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: Percentage point difference is measured as the percent of observations identified as a particular race/ethnicity in Supplemental Race minus the percent in TAF Race. Therefore, a positive value denotes an increase in the race/ethnic category when using the Supplemental Race approach. Kansas, Nebraska, Rhode Island, and Tennessee reported no White beneficiaries in TAF and started at 0% for this comparison map. DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

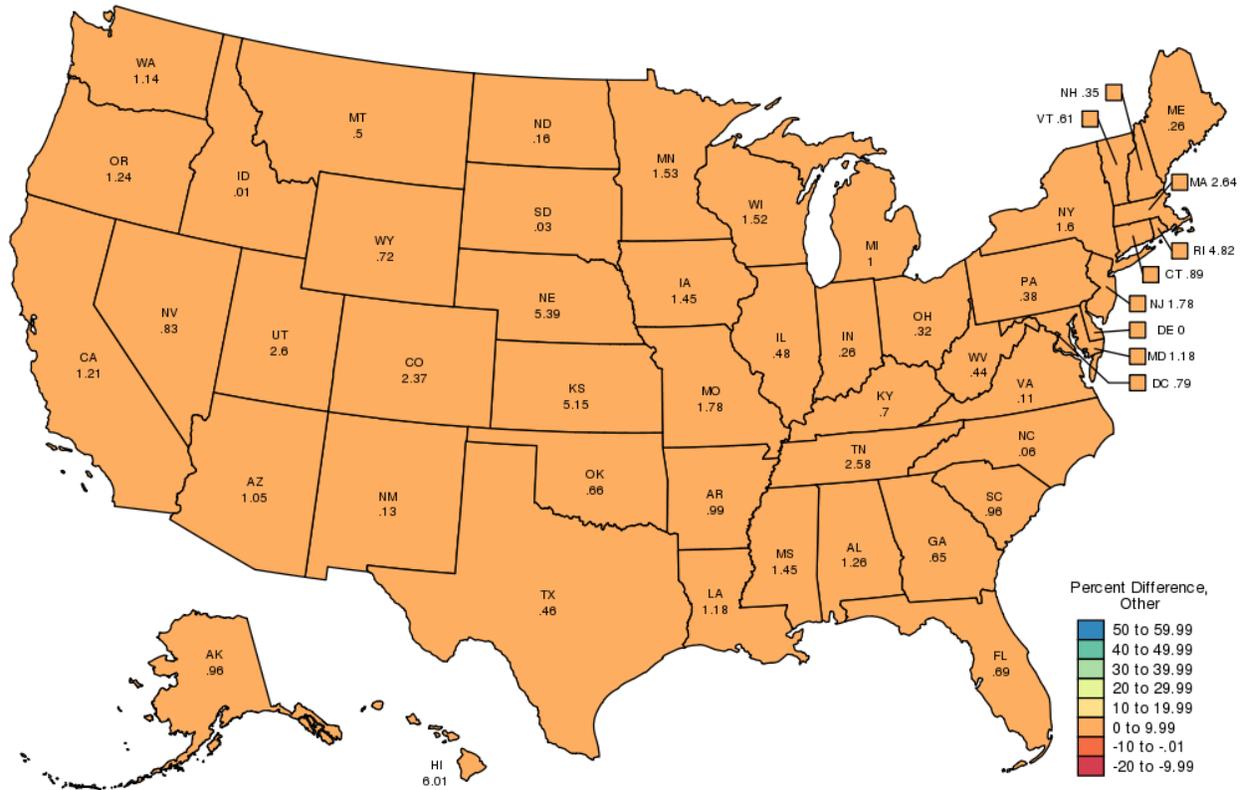
**Figure B-7. Percentage Point Difference in Race/Ethnicity Distribution Between TAF Race and Supplemental Race, Black**



Source: T-MSIS Analytic Files (TAF) data (2016); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: Percentage point difference is measured as the percent of observations identified as a particular race/ethnicity in Supplemental Race minus the percent in TAF Race. Therefore, a positive value denotes an increase in the race/ethnic category when using the Supplemental Race approach. Kansas, Nebraska, Rhode Island, and Tennessee reported no Black beneficiaries in TAF and started at 0% for this comparison map. DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

**Figure B-8. Percentage Point Difference in Race/Ethnicity Distribution Between TAF Race and Supplemental Race, Other Race**

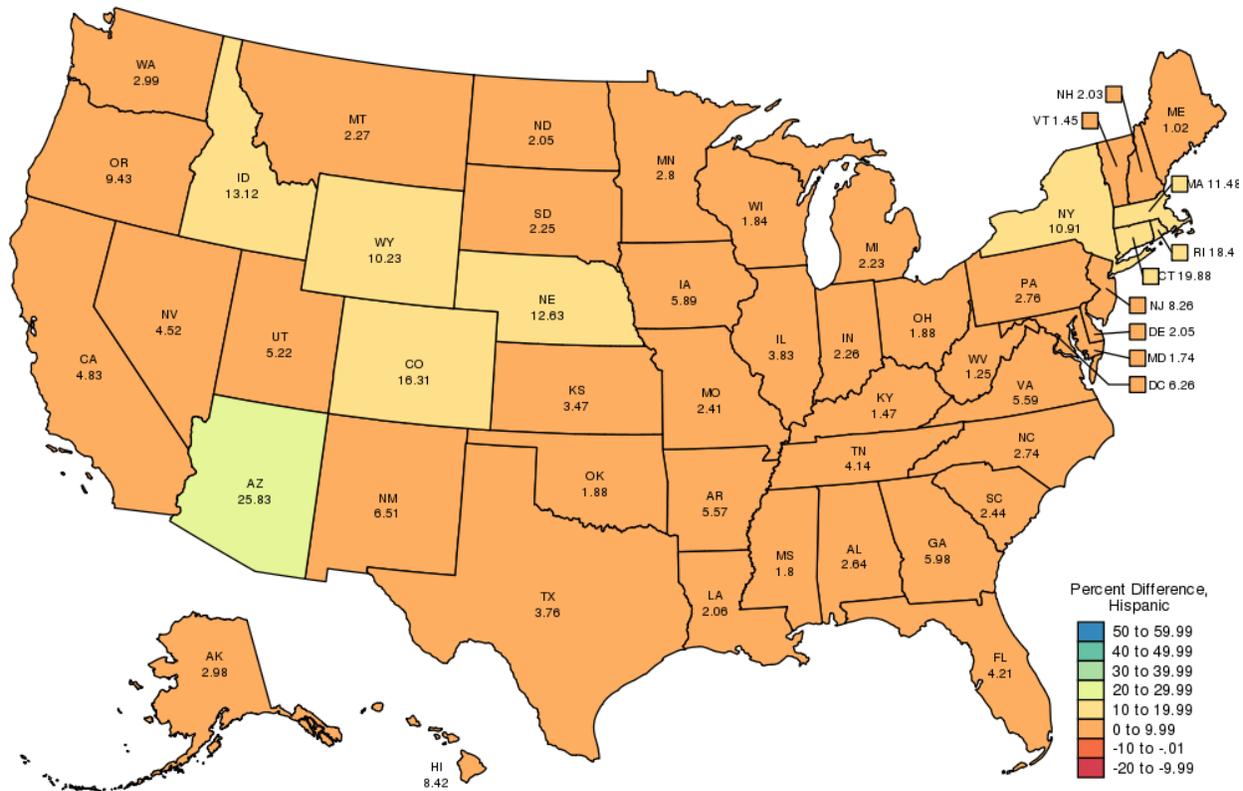


Source: T-MSIS Analytic Files (TAF) data (2016); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: Percentage point difference is measured as the percent of observations identified as a particular race/ethnicity in Supplemental Race minus the percent in TAF Race. Therefore, a positive value denotes an increase in the race/ethnic category when using the Supplemental Race approach. Nebraska, Rhode Island, and Tennessee reported no other race beneficiaries in TAF and started at 0% for this comparison map. DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

Figure B-9 through Figure B-12 provide information on the percentage point difference between TAF Race and Prioritized Race for each of the four race/ethnic groups across all states. Nebraska, Rhode Island, and Tennessee were missing entirely on race/ethnicity information in TAF data, so these states started at 0%. All maps used the same scale and color theme to make comparisons across racial/ethnic groups easier.

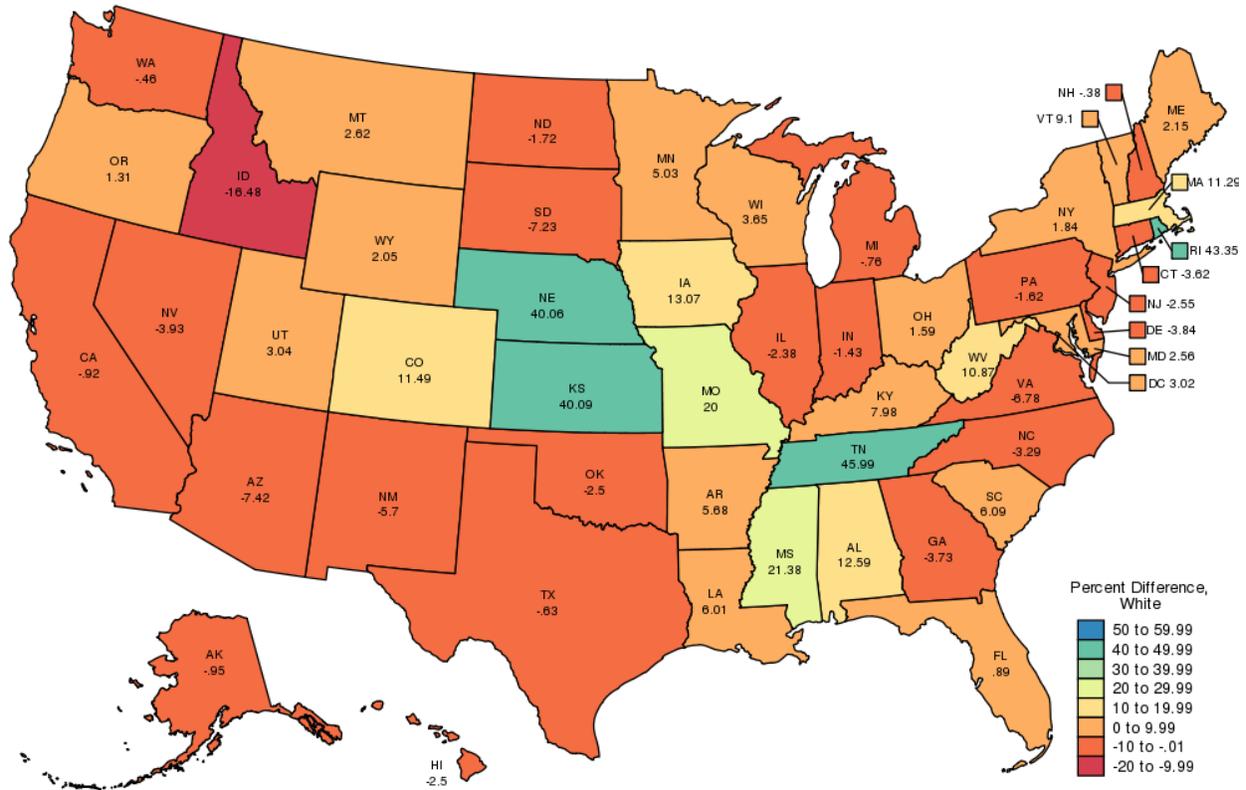
**Figure B-9. Percentage Point Difference in Race/Ethnicity Distribution Between TAF Race and Prioritized Race, Hispanic**



Source: T-MSIS Analytic Files (TAF) data (2016); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Note: Percentage point difference is measured as the percent of observations identified as a particular race/ethnicity according to the Prioritized Race approach minus the percent of individuals of this race/ethnicity according to the TAF Race. Therefore, a positive value denotes an increase in the race/ethnicity category when using the Prioritized Race approach. Connecticut and DC reported no Hispanic beneficiaries in TAF and started at 0% for this comparison map. DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

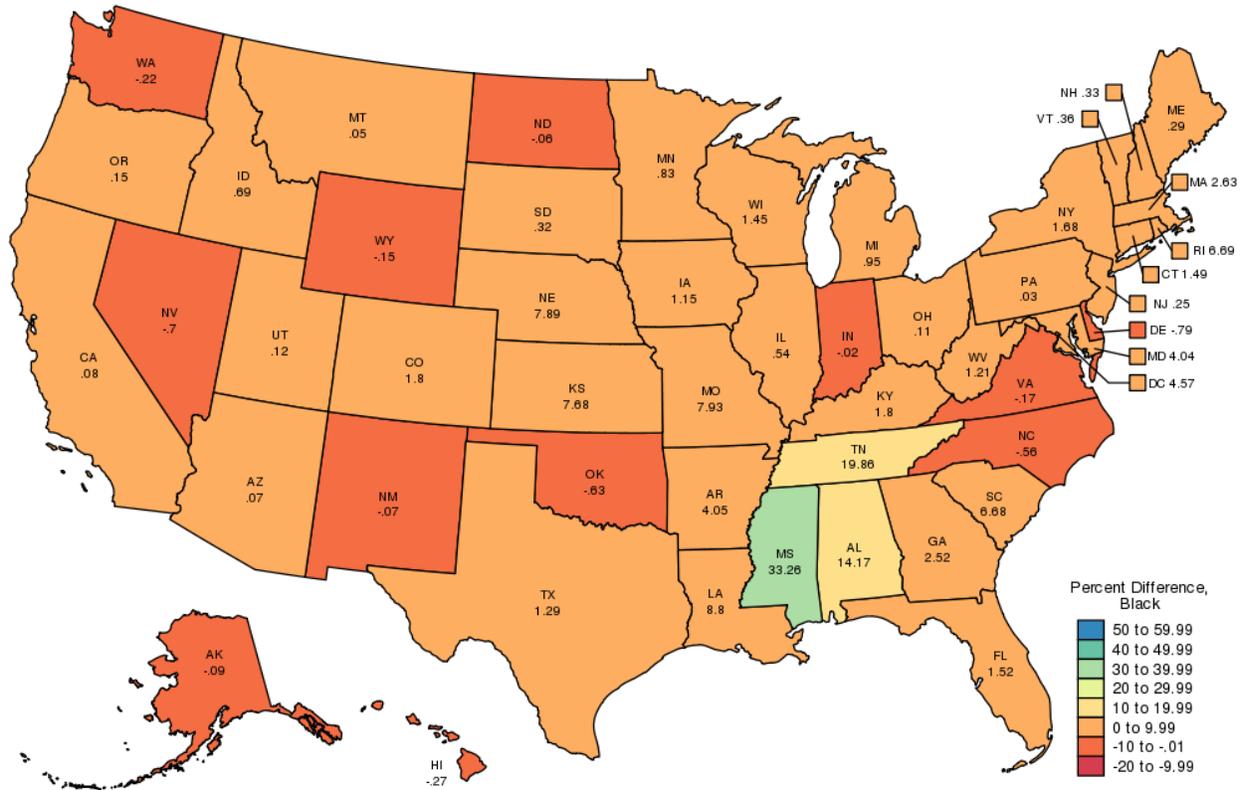
**Figure B-10. Percentage Point Difference in Race/Ethnicity Distribution Between TAF Race and Prioritized Race, White**



Source: T-MSIS Analytic Files (TAF) data (2016); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Note: Percentage point difference is measured as the percent of observations identified as a particular race/ethnicity according to the Prioritized Race approach minus the percent of individuals of this race/ethnicity according to the TAF Race. Therefore, a positive value denotes an increase in the race/ethnicity category when using the Prioritized Race approach. Kansas, Nebraska, Rhode Island, and Tennessee reported no White beneficiaries in TAF and started at 0% for this comparison map. DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

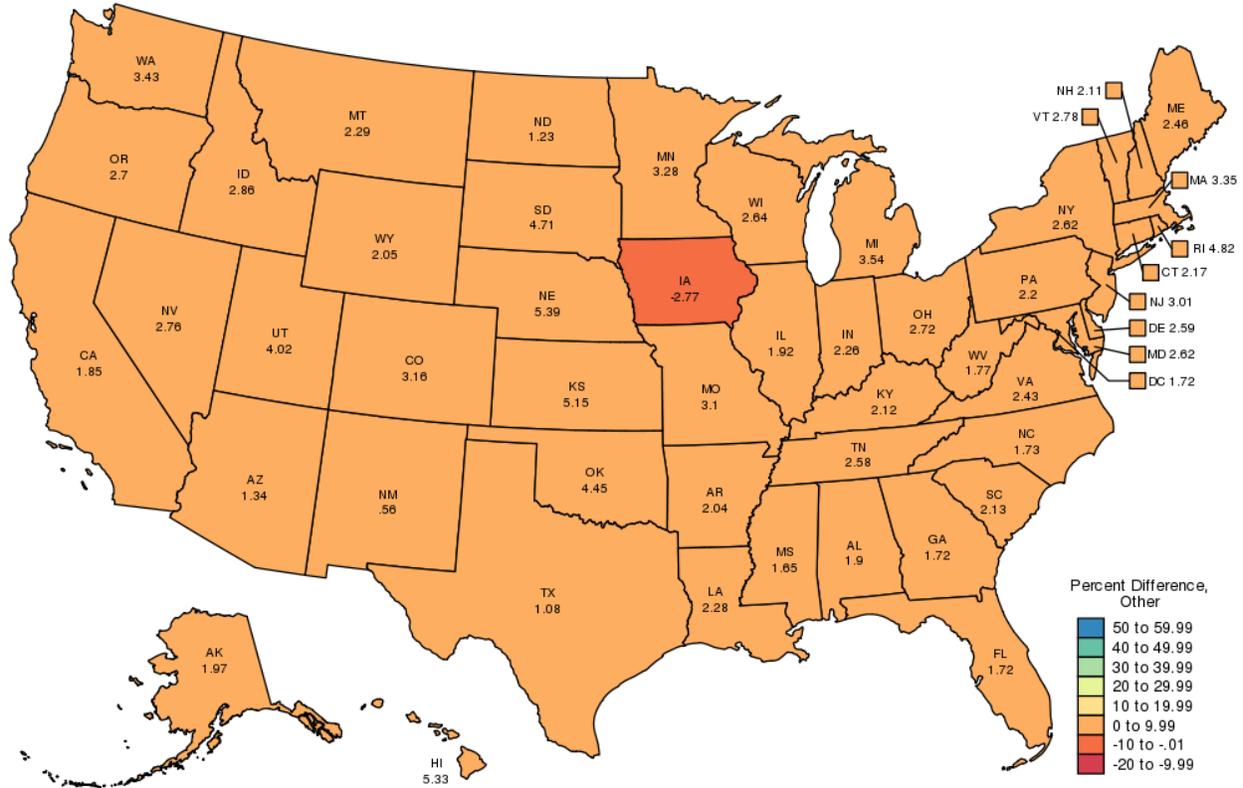
**Figure B-11. Percentage Point Difference in Race/Ethnicity Distribution Between TAF Race and Prioritized Race, Black**



Source: T-MSIS Analytic Files (TAF) data (2016); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Note: Percentage point difference is measured as the percent of observations identified as a particular race/ethnicity according to the Prioritized Race approach minus the percent of individuals of this race/ethnicity according to the TAF Race. Therefore, a positive value denotes an increase in the race/ethnicity category when using the Prioritized Race approach. Kansas, Nebraska, Rhode Island, and Tennessee reported no Black beneficiaries in TAF and started at 0% for this comparison map. DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

**Figure B-12. Percentage Point Difference in Race/Ethnicity Distribution Between TAF Race and Prioritized Race, Other Race**



Source: T-MSIS Analytic Files (TAF) data (2016); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Note: Percentage point difference is measured as the percent of observations identified as a particular race/ethnicity according to the Prioritized Race approach minus the percent of individuals of this race/ethnicity according to the TAF Race. Therefore, a positive value denotes an increase in the race/ethnicity category when using the Prioritized Race approach. Nebraska, Rhode Island, and Tennessee reported no other race beneficiaries in TAF and started at 0% for this comparison map. DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

Table B-2 and Table B-3 provide the state-level percentages in table format for the percent difference between TAF Race and Supplemental/Prioritized Race, respectively (Figure B-5 through Figure B-12).

**Table B-2. Percentage Point Difference between TAF Race and Supplemental Race for each of the Four Race/Ethnicity Categories**

	<b>Figure B-5</b>	<b>Figure B-6</b>	<b>Figure B-7</b>	<b>Figure B-8</b>
	Percentage point difference, TAF and Supplemental Race, Hispanic	Percentage point difference, TAF and Supplemental Race, White	Percentage point difference, TAF and Supplemental Race, Black	Percentage point difference, TAF and Supplemental Race, Other
<b>State</b>	(1)	(2)	(3)	(4)
<b>AK</b>	0.56	2.17	0.22	0.96
<b>AL</b>	2.21	13.57	14.27	1.26
<b>AR</b>	4.51	8.27	3.57	0.99
<b>AZ</b>	13.61	4.34	0.82	1.05
<b>CA</b>	2.00	2.09	0.55	1.21
<b>CO</b>	14.12	14.27	1.99	2.37
<b>CT</b>	11.40	5.22	2.41	0.89
<b>DC</b>	5.34	3.19	6.25	0.79
<b>DE</b>	0.00	0.00	0.00	0.00
<b>FL</b>	2.70	2.80	2.14	0.69
<b>GA</b>	0.62	2.18	3.04	0.65
<b>HI</b>	1.41	3.34	0.22	6.01
<b>IA</b>	1.17	13.13	1.60	1.45
<b>ID</b>	0.02	0.17	0.00	0.01
<b>IL</b>	0.95	1.63	0.86	0.48
<b>IN</b>	0.59	1.74	0.47	0.26
<b>KS</b>	3.47	40.09	7.68	5.15
<b>KY</b>	0.60	10.34	1.73	0.70
<b>LA</b>	1.05	8.06	8.86	1.18
<b>MA</b>	9.85	13.24	3.03	2.64
<b>MD</b>	1.01	4.02	4.76	1.18
<b>ME</b>	0.14	5.37	0.16	0.26
<b>MI</b>	0.36	3.14	1.46	1.00
<b>MN</b>	1.52	6.89	2.00	1.53
<b>MO</b>	1.75	21.88	8.04	1.78
<b>MS</b>	1.64	21.68	33.33	1.45
<b>MT</b>	0.40	6.24	0.09	0.50
<b>NC</b>	0.13	0.28	0.16	0.06
<b>ND</b>	0.05	1.27	0.03	0.16
<b>NE</b>	12.63	40.06	7.89	5.39
<b>NH</b>	0.32	3.25	0.17	0.35
<b>NJ</b>	2.16	3.82	1.20	1.78
<b>NM</b>	0.40	0.73	0.05	0.13
<b>NV</b>	0.96	0.57	0.28	0.83
<b>NY</b>	8.35	4.62	2.47	1.60
<b>OH</b>	0.33	4.82	0.83	0.32

	<b>Figure B-5</b>	<b>Figure B-6</b>	<b>Figure B-7</b>	<b>Figure B-8</b>
	Percentage point difference, TAF and Supplemental Race, Hispanic	Percentage point difference, TAF and Supplemental Race, White	Percentage point difference, TAF and Supplemental Race, Black	Percentage point difference, TAF and Supplemental Race, Other
<b>State</b>	(1)	(2)	(3)	(4)
<b>OK</b>	0.27	1.75	0.51	0.66
<b>OR</b>	6.86	5.22	0.29	1.24
<b>PA</b>	0.88	1.63	0.47	0.38
<b>RI</b>	18.40	43.35	6.69	4.82
<b>SC</b>	1.40	8.09	6.89	0.96
<b>SD</b>	0.00	0.01	0.01	0.03
<b>TN</b>	4.14	45.99	19.86	2.58
<b>TX</b>	1.44	1.89	1.71	0.46
<b>UT</b>	1.75	7.77	0.27	2.60
<b>VA</b>	0.43	0.34	0.20	0.11
<b>VT</b>	0.28	12.59	0.20	0.61
<b>WA</b>	1.17	3.10	0.33	1.14
<b>WI</b>	0.71	5.46	1.90	1.52
<b>WV</b>	0.33	13.51	0.82	0.44
<b>WY</b>	5.96	7.34	0.15	0.72

Source: T-MSIS Analytic Files (TAF) data (2016); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

**Table B-3. Percentage Point Difference between TAF Race and Prioritized Race for each of the Four Race/Ethnicity Categories**

	<b>Figure B-9</b>	<b>Figure B-10</b>	<b>Figure B-11</b>	<b>Figure B-12</b>
	Percentage point difference, TAF and Prioritized Race, Hispanic	Percentage point difference, TAF and Prioritized Race, White	Percentage point difference, TAF and Prioritized Race, Black	Percentage point difference, TAF and Prioritized Race, Other
<b>State</b>	(1)	(2)	(3)	(4)
<b>AK</b>	2.98	-0.95	-0.09	1.97
<b>AL</b>	2.64	12.59	14.17	1.90
<b>AR</b>	5.57	5.68	4.05	2.04
<b>AZ</b>	25.83	-7.42	0.07	1.34
<b>CA</b>	4.83	-0.92	0.08	1.85
<b>CO</b>	16.31	11.49	1.80	3.16
<b>CT</b>	19.88	-3.62	1.49	2.17
<b>DC</b>	6.26	3.02	4.57	1.72
<b>DE</b>	2.05	-3.84	-0.79	2.59
<b>FL</b>	4.21	0.89	1.52	1.72
<b>GA</b>	5.98	-3.73	2.52	1.72
<b>HI</b>	8.42	-2.50	-0.27	5.33
<b>IA</b>	5.89	13.07	1.15	-2.77
<b>ID</b>	13.12	-16.48	0.69	2.86

	<b>Figure B-9</b>	<b>Figure B-10</b>	<b>Figure B-11</b>	<b>Figure B-12</b>
	Percentage point difference, TAF and Prioritized Race, Hispanic	Percentage point difference, TAF and Prioritized Race, White	Percentage point difference, TAF and Prioritized Race, Black	Percentage point difference, TAF and Prioritized Race, Other
<b>State</b>	(1)	(2)	(3)	(4)
<b>IL</b>	3.83	-2.38	0.54	1.92
<b>IN</b>	2.26	-1.43	-0.02	2.26
<b>KS</b>	3.47	40.09	7.68	5.15
<b>KY</b>	1.47	7.98	1.80	2.12
<b>LA</b>	2.06	6.01	8.80	2.28
<b>MA</b>	11.48	11.29	2.63	3.35
<b>MD</b>	1.74	2.56	4.04	2.62
<b>ME</b>	1.02	2.15	0.29	2.46
<b>MI</b>	2.23	-0.76	0.95	3.54
<b>MN</b>	2.80	5.03	0.83	3.28
<b>MO</b>	2.41	20.00	7.93	3.10
<b>MS</b>	1.80	21.38	33.26	1.65
<b>MT</b>	2.27	2.62	0.05	2.29
<b>NC</b>	2.74	-3.29	-0.56	1.73
<b>ND</b>	2.05	-1.72	-0.06	1.23
<b>NE</b>	12.63	40.06	7.89	5.39
<b>NH</b>	2.03	-0.38	0.33	2.11
<b>NJ</b>	8.26	-2.55	0.25	3.01
<b>NM</b>	6.51	-5.70	-0.07	0.56
<b>NV</b>	4.52	-3.93	-0.70	2.76
<b>NY</b>	10.91	1.84	1.68	2.62
<b>OH</b>	1.88	1.59	0.11	2.72
<b>OK</b>	1.88	-2.50	-0.63	4.45
<b>OR</b>	9.43	1.31	0.15	2.70
<b>PA</b>	2.76	-1.62	0.03	2.20
<b>RI</b>	18.40	43.35	6.69	4.82
<b>SC</b>	2.44	6.09	6.68	2.13
<b>SD</b>	2.25	-7.23	0.32	4.71
<b>TN</b>	4.14	45.99	19.86	2.58
<b>TX</b>	3.76	-0.63	1.29	1.08
<b>UT</b>	5.22	3.04	0.12	4.02
<b>VA</b>	5.59	-6.78	-0.17	2.43
<b>VT</b>	1.45	9.10	0.36	2.78
<b>WA</b>	2.99	-0.46	-0.22	3.43
<b>WI</b>	1.84	3.65	1.45	2.64
<b>WV</b>	1.25	10.87	1.21	1.77
<b>WY</b>	10.23	2.05	-0.15	0.00

Source: T-MSIS Analytic Files (TAF) data (2016); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

Table B-4 is similar to Table 8 but provides information on the race/ethnicity distribution for the detailed race/ethnicity variable. There was very little change in the percentage of Asian, AIAN, and NHPI beneficiaries across the three different race/ethnicity variable approaches. However, there was an increase in the percent of beneficiaries identifying as two or more races in both the Supplemental and Prioritized Race approaches.

**Table B-4. Race/Ethnicity Distribution in TAF Data Before and After Linking to Decennial/ACS data**

	TAF Race		Supplemental Race		Prioritized Race	
	N	Column %	N	Column %	N	Column %
	(1)	(2)	(3)	(4)	(5)	(6)
<b>Hispanic</b>	21,120,000	27.32	24,020,000	26.94	26,580,000	29.81
<b>White</b>	34,060,000	44.06	39,530,000	44.34	37,110,000	41.62
<b>Black</b>	16,190,000	20.94	18,620,000	20.88	18,640,000	20.91
<b>Asian</b>	4,107,000	5.31	4,661,000	5.23	4,812,000	5.40
<b>AIAN</b>	1,179,000	1.53	1,269,000	1.42	1,151,000	1.29
<b>NHPI</b>	480,000	0.62	508,000	0.57	349,000	0.39
<b>Two or more races</b>	163,000	0.21	555,000	0.62	519,000	0.58
<b>Total</b>	77,300,000	100	89,160,000	100	89,160,000	100

Source: T-MSIS Analytic Files (TAF) data (2016); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDREB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy

Appendix C. Bias Assessment Supplement

Table C-1 is similar to Table 9 but provides information on the race/ethnicity distribution for the detailed race/ethnicity variable. Like our findings for Hispanic, White, and Black beneficiaries missing TAF race/ethnicity, Asian, AIAN, NHPI, and beneficiaries identifying as two or more races who were missing TAF race/ethnicity had higher rates of mortality relative to those who were not missing this information in TAF data.

**Table C-1. Average Annual Mortality Rate per 10,000 Beneficiaries by Missing Status and Time Period across Detailed Race/Ethnicity Groups**

	Before the Pandemic (January 2016 - February 2020)		Including the Pandemic (January 2016 - June 2021)		N	
	Not missing TAF race; TAF race	Missing TAF race; Decennial/ACS race	Not missing TAF race; TAF race	Missing TAF race; Decennial/ACS race	Not missing TAF race; TAF race	Missing TAF race; Decennial/ACS race
<b>Hispanic</b>	42.08	49.80	46.13	55.77	21,120,000	2,901,000
<b>White</b>	131.60	160.40	130.20	157.90	34,060,000	5,466,000
<b>Black</b>	84.92	92.49	89.24	99.77	16,190,000	2,435,000
<b>Asian</b>	67.55	89.93	72.75	95.17	4,107,000	553,700
<b>AIAN</b>	75.96	138.10	83.72	141.20	1,179,000	89,540
<b>NHPI</b>	65.51	83.63	68.49	85.38	479,600	28,240
<b>Two or more races</b>	35.10	44.56	37.13	46.59	163,400	391,500

Source: T-MSIS Analytic Files (TAF) data (2016); Census Numident (Q2, 2021); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy

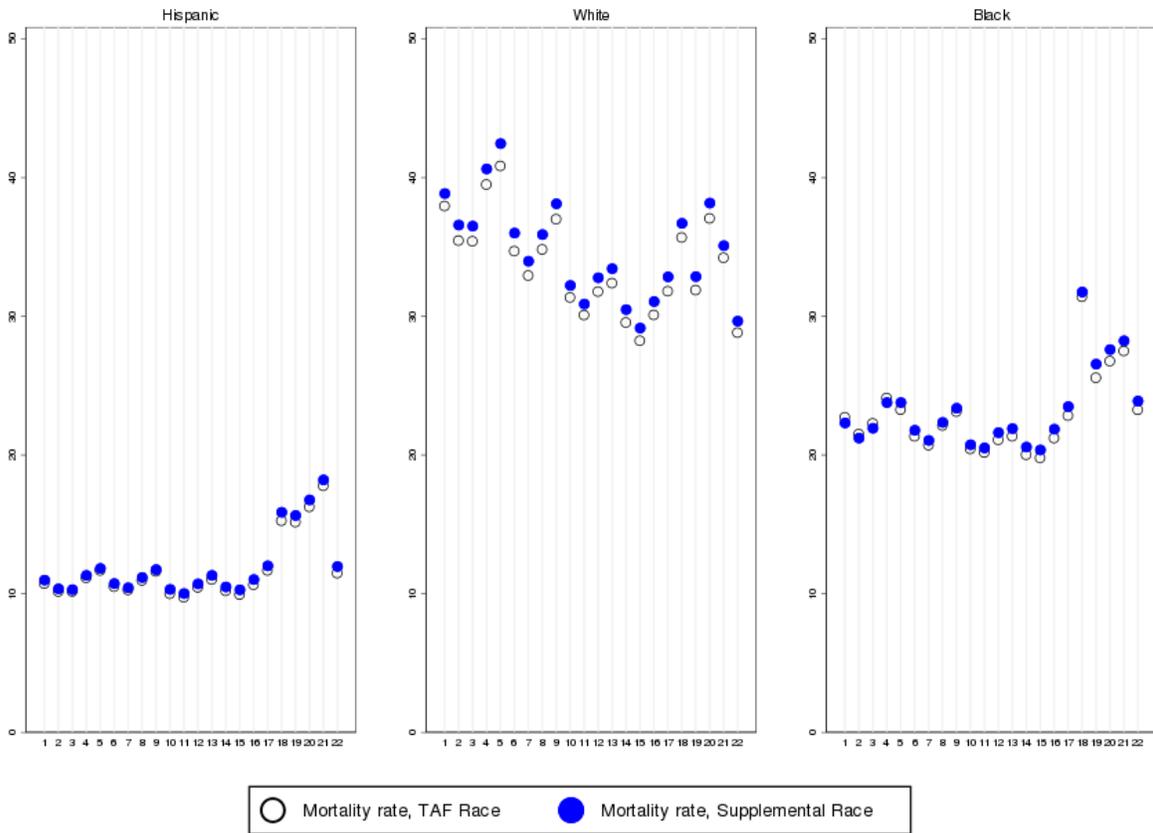
Figure C-1 and Figure C-2 provide national race-specific quarterly mortality rates per 10,000 for Hispanic, White, and Black Medicaid beneficiaries for the entire mortality period (January 2016 to June 2021). Quarterly mortality rates were estimated for these national figures since the goal was to demonstrate how bias in mortality rates evolved over time and annual mortality rates masked important trends. We calculated a quarterly all-cause mortality rate per 10,000 beneficiaries for each demographic group. Specifically, we divided the number of beneficiaries who died in each quarter by the total number of beneficiaries alive at the start of that quarter. Mortality rates for TAF Race (hollow dots) and Supplemental Race (blue dots) are presented in Figure C-1, while mortality rates for TAF Race (hollow dots) and Prioritized Race (green dots) are presented in Figure C-2. In both figures, there is a pandemic jump in mortality rate for all three groups at quarter 18 (April 2020 – June 2020), which covered the first pandemic wave in the U.S. We see higher mortality rates for White beneficiaries since these are age-unadjusted mortality rates and this group is much older relative to other groups.

For the purposes of this analysis, potential bias is the difference between the hollow dots (representing mortality rates calculated for Medicaid beneficiaries with race/ethnicity information in TAF data) and the blue dots (representing mortality rates calculated for Medicaid beneficiaries where we add beneficiaries whose race/ethnicity we found in Decennial/ACS data). Figure C-1 shows that mortality rates are slightly underestimated for all three racial/ethnic groups when beneficiaries with missing race/ethnicity in TAF data are omitted (TAF Race) for most quarters. Estimates for White mortality had the largest discrepancy once missing information was filled in with Decennial/ACS data. For both Figure C-1 and Figure C-2, there is no noticeable trend in mortality rate bias over this period.<sup>18</sup>

---

<sup>18</sup> We caution the reader when interpreting any mortality rate trends observed in these two figures. It appears that mortality rate for White individuals is trending downward during this period. This decline is a function of analytic approach and how we calculated these mortality rates. Quarterly mortality rates were calculated using the number of people who died in a given quarter divided by the number of people alive at the beginning of each quarter. However, the population is fixed in that we are focusing exclusively on individuals who were eligible for Medicaid at some point in 2016. Therefore, the observed decline in mortality reflects the fixed population. That is, the oldest and sickest in our fixed Medicaid population are more likely to die sooner, thus, leaving a younger and possibly healthier group of beneficiaries. The mortality rate is expected to decline over time. Our goal was not to assess changes in mortality rates for this group over time but rather to evaluate how bias was potentially introduced into the mortality rate excluding beneficiaries with unknown race/ethnicity in TAF data.

**Figure C-1. Quarterly Race-Specific Mortality Rates per 10,000 Medicaid Beneficiaries, TAF Race and Supplemental Race Across All States, Including the Pandemic (January 2016 – June 2021)**

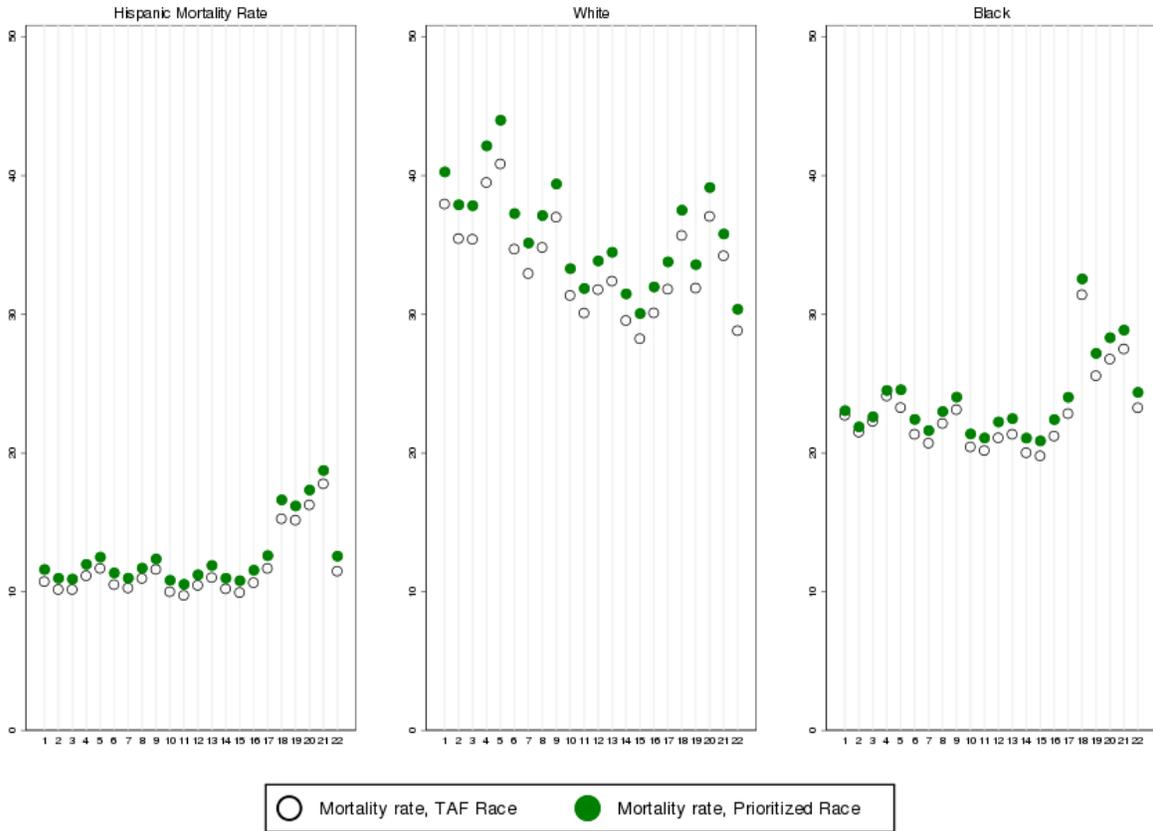


Source: T-MSIS Analytic Files (TAF) data (2016); Census Numident (Q2, 2021); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

When comparing TAF Race and Prioritized Race in Figure C-2, there is a noticeable increase in the mortality rate bias relative to the observed bias in Figure C-1. The difference in mortality rates is especially noticeable for White beneficiaries. We find that, at the national level, race-specific mortality rates are underestimated when compared to rates that take full advantage of Decennial/ACS data. Overall, we find some evidence that mortality rate may be underestimated when relying exclusively on TAF data. The degree and severity of this bias vary across racial/ethnic groups and depend on which approach we use when utilizing Decennial/ACS data (i.e., Supplemental or Prioritized Race). Mortality rates presented here are estimated across all states. Given the state-level variation in missing race/ethnicity information in Medicaid data, we anticipate the magnitude of the bias to vary across states. Next, we look at estimates of mortality rate across states.

**Figure C-2. Quarterly Race-Specific Mortality Rates per 10,000 Medicaid Beneficiaries, TAF Race and Prioritized Race Across All states, Including the Pandemic (January 2016 – June 2021)**

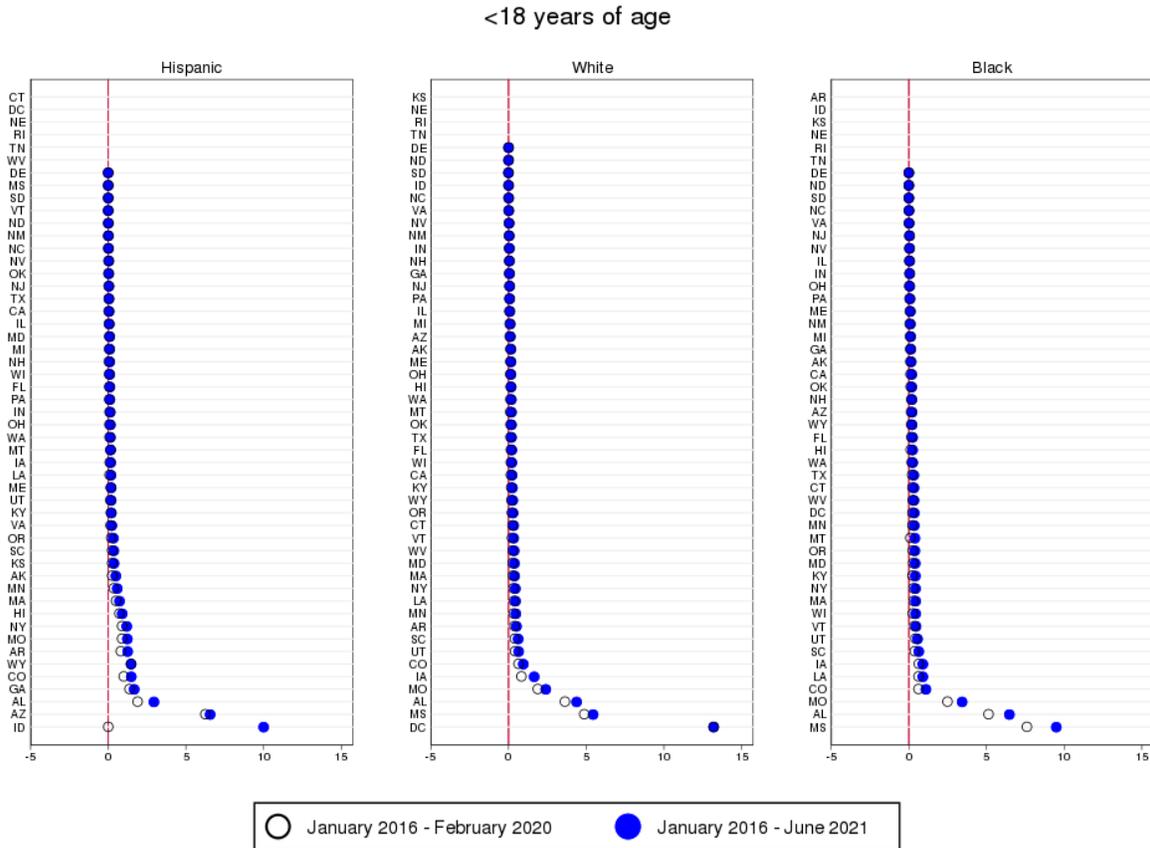


Source: T-MSIS Analytic Files (TAF) data (2016); Census Numident (Q2, 2021); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

Figure C-3 through Figure C-8 provide state-level mortality bias rates for each of the two time periods stratified by our three age-categories: <18, 18-64, and >65. There were important differences in the estimation of mortality rate biases across age groups, with older adults (>65 years of age) having the largest absolute mortality rate bias of any of the three age groups.

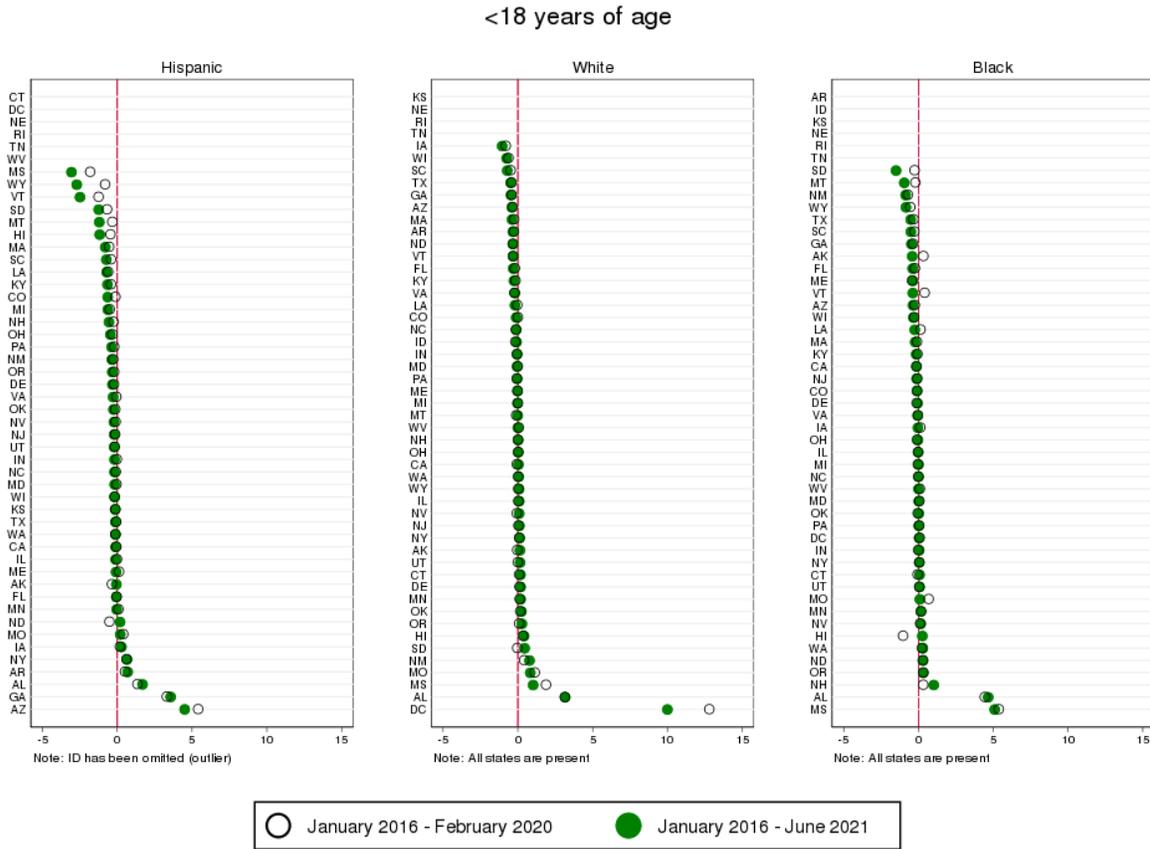
**Figure C-3. Differences in Average Annual Race-Specific Mortality per 10,000 between TAF Race and Supplemental Race for Two Time Periods for Individuals <18 Years of Age**



Source: T-MSIS Analytic Files (TAF) data (2016); Census Numident (Q2, 2021); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: Differences are calculated as the annual mortality rate in TAF Race minus the annual mortality rate in Supplemental Race. Therefore, positive values suggest that mortality rates were being overestimated and negative values suggest that mortality rates were being underestimated when relying on TAF Race. DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

**Figure C-4. Differences in Average Annual Race-Specific Mortality per 10,000 between TAF Race and Prioritized Race for Two Time Periods for Individuals <18 Years of Age**

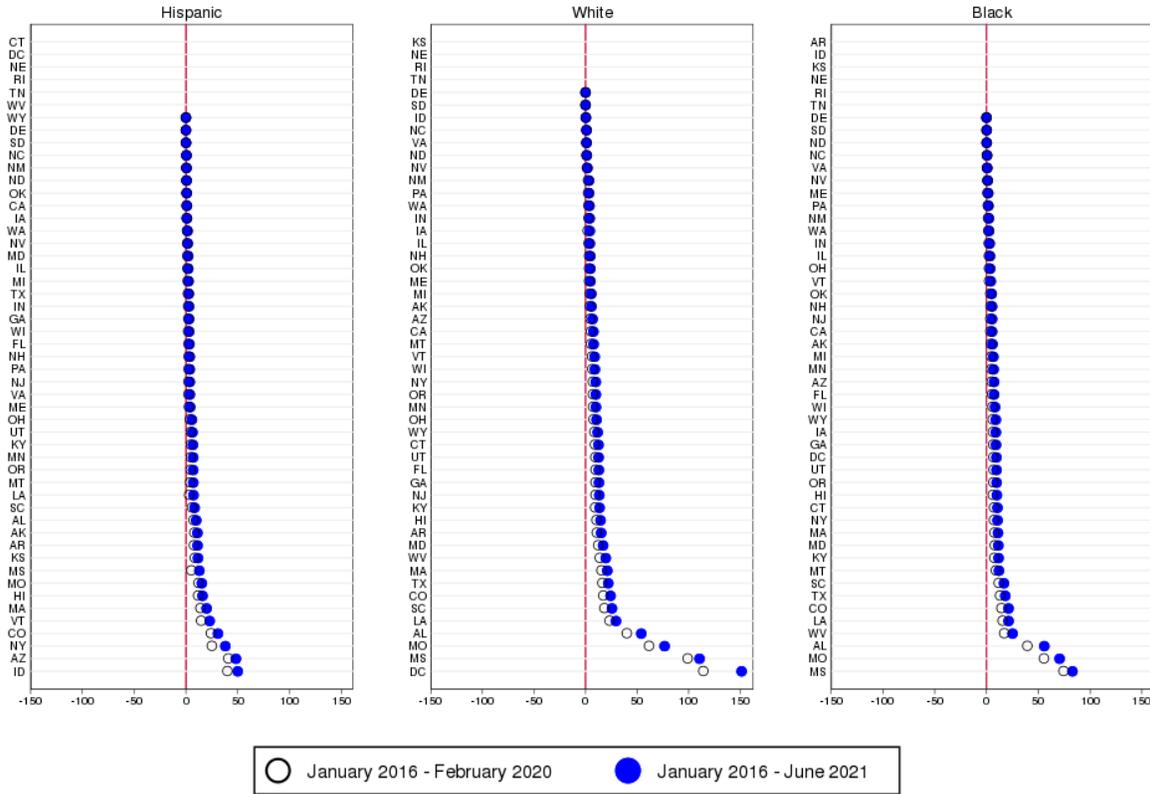


Source: T-MSIS Analytic Files (TAF) data (2016); Census Numident (Q2, 2021); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: Omitted values include ID for Hispanic (-1.99 and 26.48) beneficiaries (before the pandemic and including the pandemic, respectively). Differences are calculated as the annual mortality rate in TAF Race minus the annual mortality rate in Prioritized Race. Therefore, positive values suggest that mortality rates were being overestimated and negative values suggest that mortality rates were being underestimated when relying on TAF Race. DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

**Figure C-5. Differences in Average Annual Race-Specific Mortality per 10,000 between TAF Race and Supplemental Race for Two Time Periods for Individuals 18-64 Years of Age**

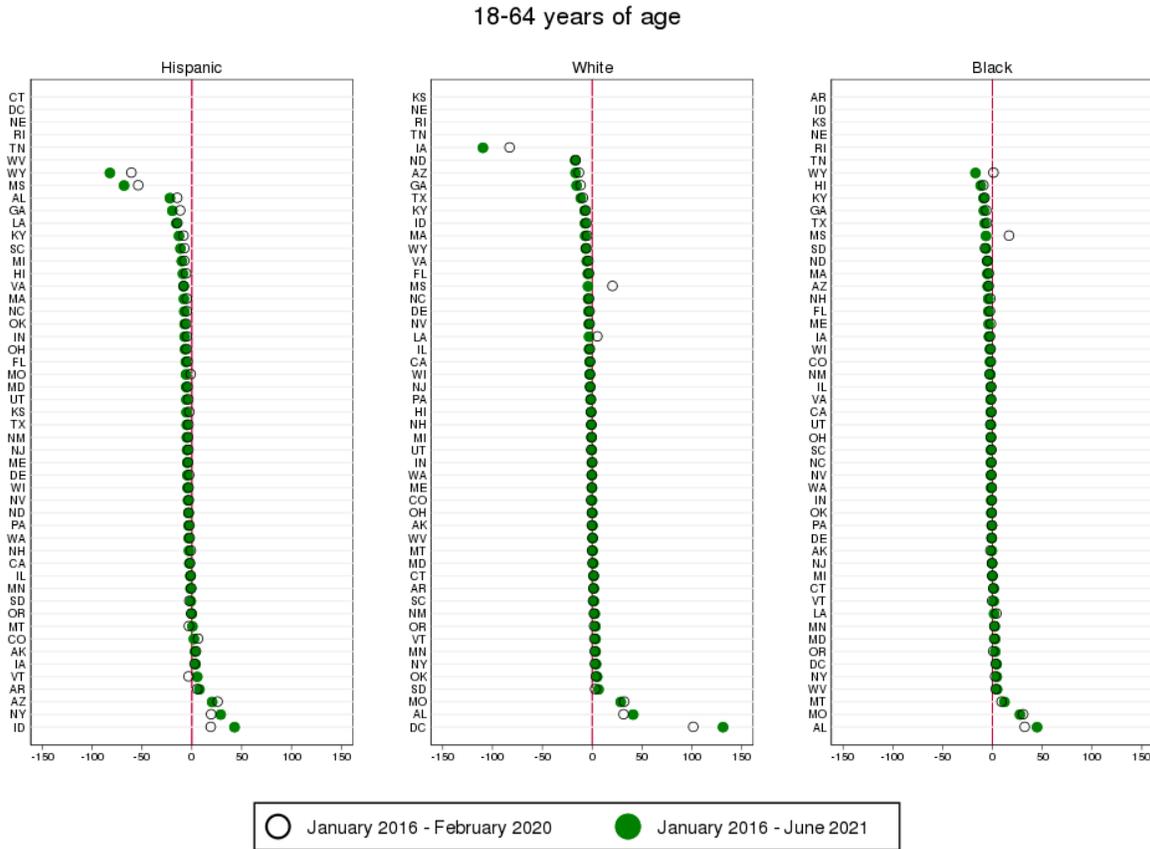
18-64 years of age



Source: T-MSIS Analytic Files (TAF) data (2016); Census Numident (Q2, 2021); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: Differences are calculated as the annual mortality rate in TAF Race minus the annual mortality rate in Supplemental Race. Therefore, positive values suggest that mortality rates were being overestimated and negative values suggest that mortality rates were being underestimated when relying on TAF Race. DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

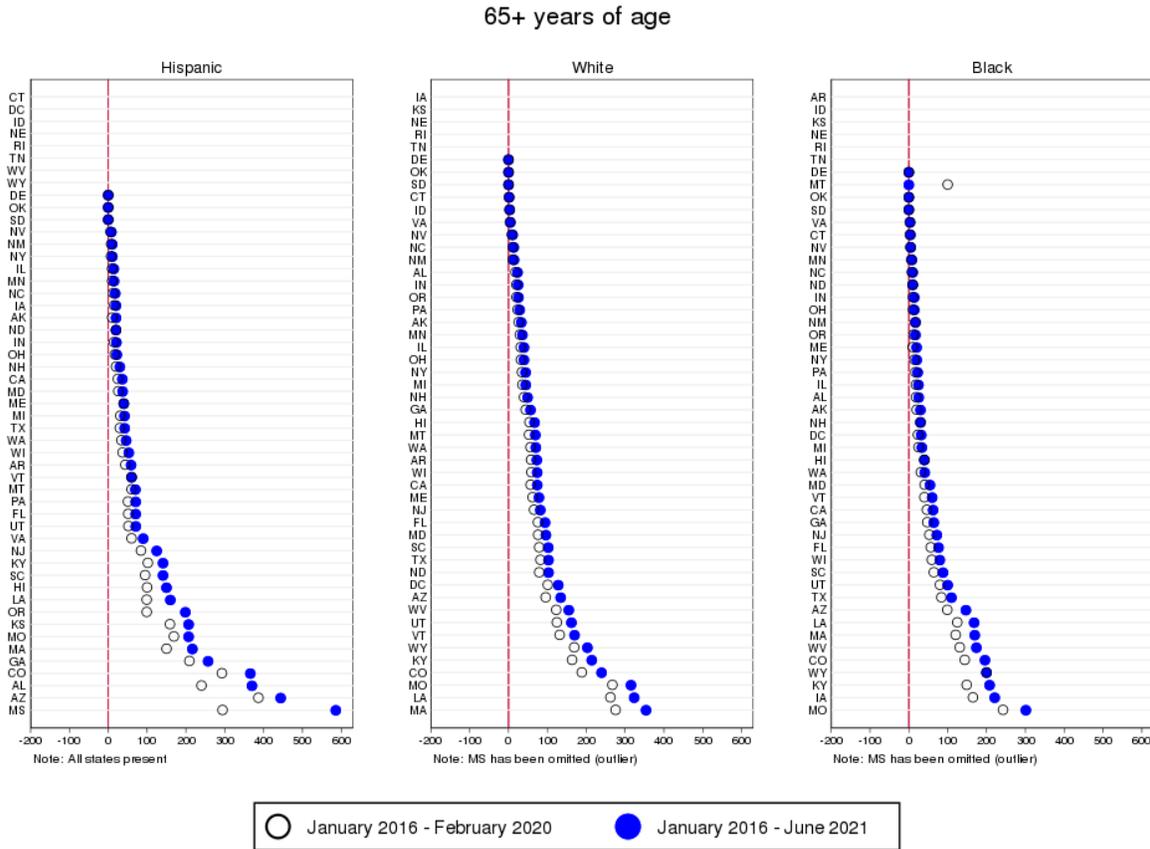
**Figure C-6. Differences in Average Annual Race-Specific Mortality per 10,000 between TAF Race and Prioritized Race for Two Time Periods for Individuals 18-64 Years of Age**



Source: T-MSIS Analytic Files (TAF) data (2016); Census Numident (Q2, 2021); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: Differences are calculated as the annual mortality rate in TAF Race minus the annual mortality rate in Prioritized Race. Therefore, positive values suggest that mortality rates were being overestimated and negative values suggest that mortality rates were being underestimated when relying on TAF Race. DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

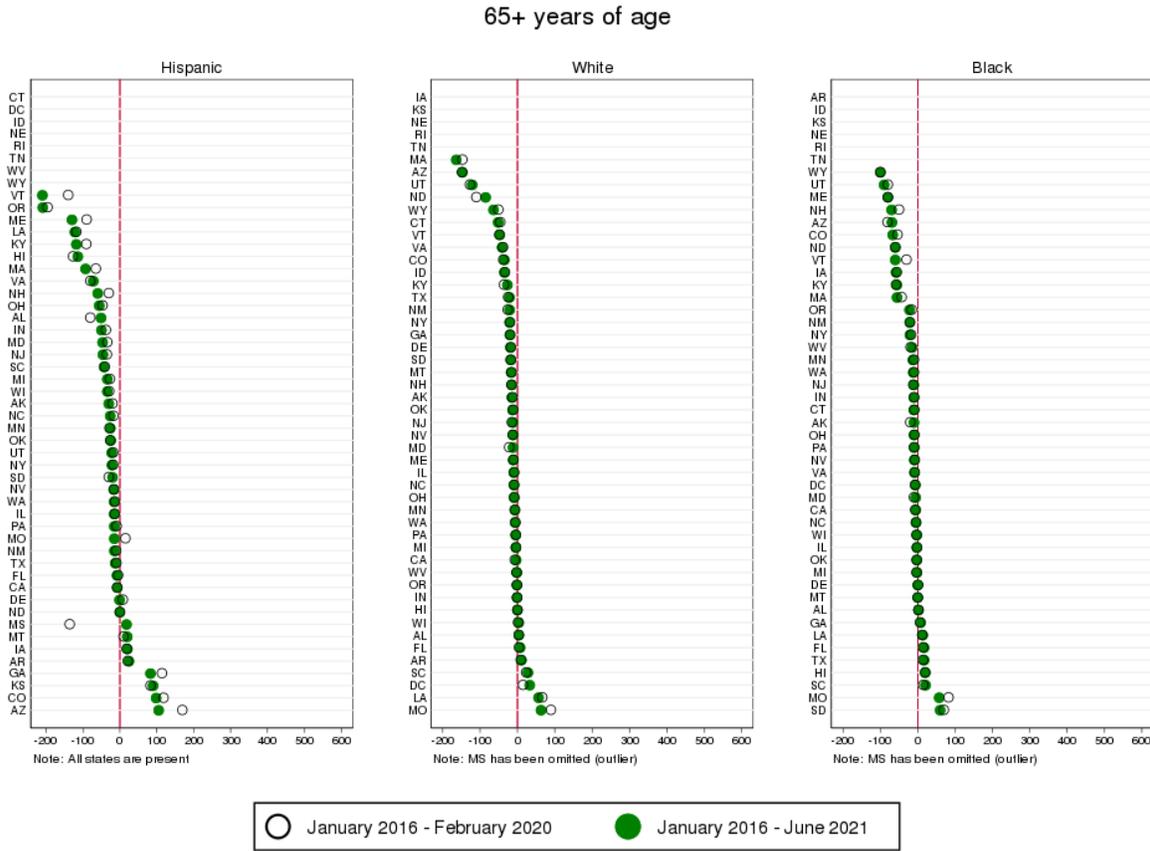
**Figure C-7. Differences in Average Annual Race-Specific Mortality per 10,000 between TAF Race and Supplemental Race for Two Time Periods for Individuals 65 years of age and older**



Source: T-MSIS Analytic Files (TAF) data (2016); Census Numident (Q2, 2021); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: Omitted values include MS for White (1203.3 and 1254.3) and Black (1078.9 and 1148.6) beneficiaries (before the pandemic and including the pandemic, respectively). Differences are calculated as the annual mortality rate in TAF Race minus the annual mortality rate in Supplemental Race. Therefore, positive values suggest that mortality rates were being overestimated and negative values suggest that mortality rates were being underestimated when relying on TAF Race. DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

**Figure C-8. Differences in Average Annual Race-Specific Mortality per 10,000 between TAF Race and Prioritized Race for Two Time Periods for Individuals 65 years of age and older**



Source: T-MSIS Analytic Files (TAF) data (2016); Census Numident (Q2, 2021); Decennial Census (2000, 2010); and American Community Survey (ACS) data (2001-2019)

Notes: Omitted values include MS for White (743 and 618) and Black (724.9 and 625.6) beneficiaries (before the pandemic and including the pandemic, respectively). Differences are calculated as the annual mortality rate in TAF Race minus the annual mortality rate in Prioritized Race. Therefore, positive values suggest that mortality rates were being overestimated and negative values suggest that mortality rates were being underestimated when relying on TAF Race. DMS #P-7522052. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY22-375. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.